nature research

# Peer Review Information

**Journal:** Nature Genetics
**Manuscript Title:** Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic, and molecular genetic levels of analysis
**Corresponding author name(s):** Dr Andrew Grotzinger

## Reviewer Comments & Decisions:

**Decision Letter, initial version:**

11th November 2020

Dear Andrew,

Your Article entitled "Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis" has been seen by two referees, whose comments are below. While they find your work of potential interest, they have raised substantial concerns that preclude publication of the work in Nature Genetics, at least in its present form.

Should further analyses (including extensive simulations comparing the calibration and power of the proposed structured multivariate GWAS method to other multi-trait methods) allow you to fully address these criticisms, we would be willing to consider an appeal of our decision (unless, of course, something similar has by then been accepted at Nature Genetics or appeared elsewhere). This includes submission or publication of a portion of this work someplace else.

We hope you understand that until we have read the revised manuscript in its entirety we cannot promise that it would be sent back for peer review.

If you are interested in attempting to revise this manuscript for submission to Nature Genetics in the future, please contact me to discuss a potential appeal. Otherwise, we hope that you find our referees' comments helpful when preparing your manuscript for resubmission elsewhere.

Sincerely,
Kyle

Kyle Vogan, PhD
Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665

Referee expertise:

Referee #1: Genetics, psychiatric diseases, statistical methods

Referee #2: Genetics, complex traits, statistical methods

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:

Comments

This paper performs genomicSEM on 11 psychiatric traits and identified four broad factors (Neurodevelopmental, Compulsive, Psychotic, and Internalizing) that model their genetic correlations. They further checked for genetic correlation of each factor with a variety of traits, and functional enrichment of genes expressed in different tissue types for each factor. Most importantly, they looked into the genetic factors who contributions to each disorder cannot be accounted for by the factors that load onto them, and performed MR analysis to show that at least at one such locus, there is evidence of causality where its effects on alcohol abuse may account for its effects on internalizing and neurodevelopmental disorders. Given the genetic effects that cannot be attributed to the factors, and that are heterogeneous among the diseases each factor load onto, the authors have concluded there is little utility in one common factor for psychiatric disorders. My concerns lie in the use of self-report data, and whether their results are well calibrated. I have written my concerns below for the authors' considerations.

Major

1. As the authors mentioned, misdiagnoses and other non-specificity incurred through using self-reported phenotypes may affect results of the analysis. Though the authors have performed the confirmatory factor analysis removing the with self-reported cohorts, and obtained similar results (with larger difference for factor loadings on ANX, lowest congruence coefficient for internalizing factor), I would rather like to see exploratory factor analysis done without the self-report cohorts - would the same factor models be obtained, and which would be the best fit in confirmatory factor analysis?

2. How does difference in sample size between the different diseases affect the analysis - they would lead to different standard errors in genetic covariance estimates, and as explained in the original genomic SEM paper, the WLS produce a solution that is dominated by the patterns of association involving the most well-powered GWAS... maximum likelihood estimates will be most pronouced when there is lower sample overlap and the contributing univariate GWAS differ substantially in sample size.

Related to above comment, removing self-reported cohorts would reduce sample size of MDD, ANX, and ALCH significantly (>80%), and less so for ADHD (>20%) - the sample size for these diseases are larger than those of other diseases before removing self-report cohorts. Would reducing their sample size to a more comparable level as other diseases help the analysis? Given this, I have further reason to believe the exploratory step should be done again using only the cohorts without self-report.

3. Qtrait heterogeneity index: I believe this is the same index as published in the original genomic SEM paper (Q SNP)? Am I right to understand that in fitting the model where the external trait predicted the individual disorders of a given factor, its effect on the factor and the factor's effect on individual disorders are already fixed (named Step 1 in original genomic SEM paper)? If so, please write explicitly in main text and in Figure 2 and Figure S5, also indicating with arrows (perhaps of a different colour) in Figure 2 and Figure S5 which effects are fixed. At the moment, this seems not to be the case, as shown in Figure 2, where I don't see a line going from trait/SNP to factor in the "Independent Pathways" panel, indicating this effect is fixed and modelled when obtaining Chi2 for the independent effects of trait. If this effect is not modelled, then are we comparing effect of trait/SNP on disease vs effect of trait/SNP on factor? If so, a lack of difference in Chi2 between them cannot indicate anything about the relationship between disease and factor with respect to the trait/SNP (trait/SNP can have the same but independent effect on both). A difference in Chi2 can indicate the trait/SNP has different effects on factor and disease, but can't indicate anything about whether the effects on disease are mediated through the factor. Can the authors clarify? It would be great if the authors can explicitly write out how the Qtrait p values are calculated, as that would make all these confusion go away.

4. Qtrait: "Excluding significant Qtrait correlations (i.e., correlations not operating through the factor), and using the same Bonferroni correction, 17 correlations were significant for Compulsive..." what correlation is this? Genetic correlation? Authors please edit text to make things more explicit and understandable.

5. Figure 3: looking at the personality section, neuroticism is significantly heterogeneous with respect to its correlation with diseases and factor, such that there is a significant portion of its correlaiton with the diseases that is not mediated through factor - is my understanding correct? If so, which disease is this driven by? Is this possible to tell given the way Qtrait is derived? I can't tell as there's nothing about how exactly the estimation of this Qtrait is done in the methods section, please clarify. And how do we understand this result, if this is saying that there is a significant part of neuroticism-internalizing disorder correlations that cannot be explained by the internalzing factor, though the genetic correlation between neuroticism and internalizing factor is > 0.8 and from Supp Figure 7e it seems both internalising traits have similar genetic correlations with neuroticism - is that right? In my mind this is the most interesting result for internalizing disorders, but the authors make no mention in the main text or discussions. Exactly the same situation for agreeableness with neurodevelopmental disorders (though negative effect). The authors even wrote something for an opposite example in the main text "Educational attainment (EA) evinced a particular pattern of genetic associations with the individual compulsive disorders that were inconsistent with their operation via the Compulsive disorders factor, where AN was more positively associated relative to OCD and TS". I find it necessary to discuss at least these two salient observations I pointed out. It is a pity not to talk about any of this in the discussion section.

6. Accelerometer data: I am not sure what we've learnt from this. The authors also don't discuss this at all later. If the authors have insights, please add to discussions, otherwise I am unsure why this

analysis is in the paper.

7. Stratified Genomic SEM: the authors wrote that "Stratified Genomic SEM models that allowed variances of the common genetic factors, and disorder-specific effects, to vary across annotations to examine whether the degree of risk sharing and differentiation is enriched across disorders". Does this mean the method is able to identify the degrees of enrichment in variance attributable to common genetic factors, and disorder specific effects, respectively? From Figure 5 it seems the authors have obtained the enrichment for annotations for the factors identified through previous analyses rather than the individual diseases - taking this as the enrichment results for common genetic factors, where are the results for the disease specific effects? Are those estimated? If not, I can't tell what the difference of this analysis is from S-LDSC on the factors, and I would find the description misleading. Please clarify. Further, in terms of the results, as the effective sample size for each factor used in the enrichment analysis are different, the power to identify significant enrichments for each factor would be different. Given the authors are using summary statistics for all analyses and can't down-sample, please discuss how difference in power affect results of this analysis.

8. Multivariate GWAS: QQplots for all GWAS show significant inflation, and though the authors have found that the intercept from LDSC were close to 1. However, an intercept close to 1 only suggest that the inflation wasn't likely caused by population stratification, and cannot account for inflation due to all statistical errors. Mostly when inflation was due only to polygenicity and not other errors, the qqplots would be inflated at the low P values (high log P), but the qqplots shown in Supplementary figure 23 all deviate from the null even at the high P values (low log P), making me very suspicious the tests are not well calibrated. The authors plot two qqplots in Supplementary figure 23, corresponding to figure 6 in the main text - one for the factor GWAS (blue), and the other for Qtrait tests (magenta). For the Qtrait P values, how are they calculated (I already asked above), are they well calibrated, and how did the authors check for calibration of these p values? I looked into the original genomicSEM paper in addition to this paper, and didn't see any tests for calibration of Qtrait. Similarly, can the authors show that the factors capture the shared effects of triats they load onto, and do not inflate their effects? It would be hard to trust these values without demonstration they are calibrated.

Minor

1. Figure S8 has nano-sized fonts, really difficult to read, please consider breaking up into multiple pages like Figures S4, S7, S9 etc

2. I am not opposed to polar plots but think they are unnecessarily hard to read. Normal bar plots would do the jobs just as nicely.

3. For most of the supplemental figures, perhaps it's much easier to represent these as tables.

Reviewer #2:
Remarks to the Author:

The authors use genomic structural equation modeling to infer 4 factors underlying 11 psychiatric disorders and draw inferences about the genetic architecture of these 4 factors. They then conduct a GWAS of the 4 factors, identifying 152 loci including 20 novel loci.

Section identifying the 4 factors:

(1) This is only a mild extension of the results of Lee et al. 2019 Cell (ref. 13), who analyzed 8 of these psychiatric disorders and identified 3 of these factors. The authors are transparent about this. Thus, this section should be made much shorter. Confirmatory factor analyses and other subtle details can be moved entirely to the supplement. The comments about Cai et al. 2020 Nat Genet (ref. 28), though appropriate, can be moved to the Discussion.

(2) The p-factor model seems to be of low interest. It has greater complexity, and has close to zero genetic associations in the GWAS part of the paper. Either all content on the p-factor model should be moved to the supplement, or considerable additional justification is needed as to why it is important in this paper.

(3) Minor comments: The authors should include a main Table that includes all of the information in Table 1 of Lee et al. 2019 Cell (ref. 13). Also, Figure 1C and Figure 1D should be moved to the supplement. (Not clear what Figure 1D refers to, as the text does not mention a "bifactor model".) The text fragment “prior literature indicating a high-order transdiagnostic ‘p-factor’” requires a reference, perhaps ref. 6 or ref. 6-8 as cited previously.

Section on genetic correlation of the 4 factors with external traits:

(4) Qtrait is not a compelling metric, both because statistical significance is sample size dependent, and because the null hypothesis that the genetic correlation between an external trait and the 11 psychiatric disorders operates entirely through one of the factors is not a plausible null hypothesis, such that its violation is not meaningful. It might be more interesting to assess what proportion of the genetic correlation between an external trait and the 11 psychiatric disorders operates through one of the factors (or all 4 factors). Also, Figure 2 should be moved to the supplement.

(5) Some of the genetic correlations of the 4 factors with external traits are interesting (particularly for the accelerometer traits), but it is unclear whether the insights gained are more impactful than the insights gained by assessing (only) the genetic correlations of the 11 psychiatric diseases with external traits (for example, in Figure 4, results for 4 factors look similar to results for corresponding psychiatric diseases). It would be important to discuss this.

Section on genetic enrichment of the 4 factors in annotations:

(6) "Stratified Genomic SEM allows us to ask whether pleiotropic loci are enriched within particular annotations": I agree with this comment. However, why not just stratify genetic covariance by annotation, which also assesses whether pleiotropic loci are enriched within particular annotations? (The authors actually develop this approach by developing stratified multivariate S-LDSC as the first step of their Stratified Genomic SEM method, although this is not made clear until the Methods section. Also, Lu et al. 2017 Am J Hum Genet is a published method for stratifying genetic covariance by annotation that should be cited and discussed.) It is unclear whether the insights gained from the factor approach are more impactful than the insights gained by (only) stratifying genetic covariance by annotation. It would be important to discuss this. Specifically, are there any annotations for which the genetic enrichment for one of the 4 factors is significantly larger in magnitude than the genetic enrichment for any of the 11 psychiatric disorders?

(7) Minor comment: the term "genetic correlation" usually refers to a relationship that has consistent sign across the genome. The term "pleiotropic" usually refers to a relationship that may have varying sign across the genome. This paper seems to focus on relationships that have consistent sign across the genome. Thus, either the term "genetic correlation" (or covariance) should be consistently used, or at least the terminology (including the use of the term "pleiotropic") should be carefully defined/explained.

Unstructured multivariate GWAS and structured multivariate GWAS sections:

(8) The "unstructured" GWAS (which does not make use of the 4 factors) identified 184 loci including 39 novel loci. The "structured" GWAS (which does make use of the 4 factors) identified 152 loci including 20 novel loci. Why is the structured GWAS preferred? Are its results more actionable, and if so why?

(9) How do the unstructured GWAS method and the structured GWAS method compare to the considerable literature on multi-trait GWAS methods that has been published, including Nieuwboer et al. 2016 Am J Hum Genet (not cited, same last author as this paper) and Turley et al. 2018 Nat Genet (not cited) and Lee et al. 2019 Cell (ref. 13) and Grotzinger et al. 2019 Nat Hum Behav (ref. 44; same first author as this paper)? In particular, is the structured GWAS method proposed in this paper identical to the structured GWAS method proposed by ref. 44? Although ref. 44 is cited in the structured GWAS method, this is not explicitly clear.

(10) For these sections to have high methodological impact, it would be critically important to add simulations comparing the calibration and power of the proposed structured GWAS method to other multi-trait methods in simulated data, and explain (with justification) in which scenarios the method provides an increase in power.

(11) For these sections to have high biological impact, it would be critically important to add considerable biological interpretation of the novel findings. Also, it would be preferred to empirically replicate novel findings in independent data sets, as an additional validation.

(12) QSNP is not a compelling metric, because the null hypothesis that the association between a SNP and a psychiatric disorder operates entirely through one of the factors is not a plausible null hypothesis, such that its violation is not meaningful. It might be more interesting to assess what proportion of the association between a SNP and a psychiatric disorder operates through one of the factors (or all 4 factors).

(13) Minor comments: definitely state in the Results section that 4,775,763 SNPs were tested for association (intersection across 11 psychiatric disorders) and a genome-wide significance threshold of 5e-08 was used. It would be good to justify the significance threshold by citing previous references in which this rather lax threshold was used. Also, “genome-wide S-LDSC matrix”: does this refer to using a genetic covariance matrix derived from S-LDSC instead of LDSC? This should be clarified.

The last two Results sections add little to the paper, and should be moved to the supplement.

**Author Rebuttal to Initial comments**

**Reviewer #1:**
**Remarks to the Author:**
**Comments**

**This paper performs genomicSEM on 11 psychiatric traits and identified four broad factors (Neurodevelopmental, Compulsive, Psychotic, and Internalizing) that model their genetic correlations. They further checked for genetic correlation of each factor with a variety of traits, and functional enrichment of genes expressed in different tissue types for each factor. Most importantly, they looked into the genetic factors who contributions to each disorder cannot be accounted for by the factors that load onto them, and performed MR analysis to show that at least at one such locus, there is evidence of causality where its effects on alcohol abuse may account for its effects on internalizing and neurodevelopmental disorders. Given the genetic effects that cannot be attributed to the factors, and that are heterogeneous among the diseases each factor load onto, the authors have concluded there is little utility in one common factor for psychiatric disorders. My concerns lie in the use of self-report data, and whether their results are well calibrated. I have written my concerns below for the authors' considerations.**

**Major**

**1. As the authors mentioned, misdiagnoses and other non-specificity incurred through using self-reported phenotypes may affect results of the analysis. Though the authors have performed the confirmatory factor analysis removing the with self-reported cohorts, and obtained similar results (with larger difference for factor loadings on ANX, lowest congruence coefficient for internalizing factor), I would rather like to see exploratory factor analysis done without the self-report cohorts - would the same factor models be obtained, and which would be the best fit in confirmatory factor analysis?**

*The reviewer's point is well taken. As we have already re-run the confirmatory factor analyses excluding univariate GWAS studies consisting of self-report phenotypes, we are more than happy to also run exploratory factor analyses excluding self-report GWAS.*

**2. How does difference in sample size between the different diseases affect the analysis - they would lead to different standard errors in genetic covariance estimates, and as explained in the original genomic SEM paper, the WLS produce a solution that is dominated by the patterns of association involving the most well-powered GWAS... maximum likelihood estimates will be most pronouced when there is lower sample overlap and the contributing univariate GWAS differ substantially in sample size. Related to above comment, removing self-reported cohorts would reduce sample size of MDD, ANX, and ALCH significantly (>80%), and less so for ADHD (>20%) - the sample size for these diseases are larger than those of other diseases before removing self-report cohorts. Would reducing their sample size to a more comparable level as other diseases help the analysis? Given this, I have further reason to believe the exploratory step should be done again using only the cohorts without self-report.**

*As noted in response to point 1, we will perform exploratory analyses excluding self-report cohorts. In order to fully address this concern, we will also examine the concordance of effects when estimating*

*multivariate GWAS effects excluding self-report cohorts for the 153 hits from the correlated factors model. We also plan to include a more detailed explanation in the Method section of the resubmission that includes interpretive guidelines in the presence of varying sample sizes. This will include the note that WLS estimation does not necessarily produce a solution dominated by the more well-powered GWAS, as suggested by the reviewer. Rather, WLS gives more weight to producing model estimates that most closely match the genetic correlations and heritability's associated with the better powered GWAS. Given adequate sample sizes, WLS and ML converge to the same solution. In instances where the phenotypes characterized by more well-powered GWAS estimates are not highly correlated with the other phenotypes, the model will then prioritize minimizing the influence of these particular disorders on factor GWAS estimates.*

*We believe that the decision to include certain self-report cohorts in the main analysis is defensible for a number of reasons. We note in particular the large genetic correlations between the clinically diagnosed and self-report GWAS, the increased mean chi-square when meta-analyzing self-report and clinical diagnosis GWAS (Table S42), and a general trend in psychiatric genomics to include self-report cohorts in the primary GWAS studies being published. In addition, the specific self-report cohorts we include were based on informed decisions. For example, the meta-analysis between PGC Alcohol Use Disorder and UKB self-reported alcohol use is limited to self-reported problematic alcohol use (as assessed by the AUDIT-P) and not alcohol consumption (as assessed by the AUDIT-C). This is based on prior work indicating stronger genetic correlations between self-reported problematic alcohol use and alcohol dependence relative to self-reported alcohol consumption (Sanchez-Roige et al., 2019). In addition, the choice to use the broad depression phenotype from UK Biobank and self-report 23andMe MDD phenotype was based on the inclusion of both phenotypes in the most recent GWAS of MDD (Howard et al., 2019) and of the latter phenotype in the prior PGC GWAS of MDD (Wray et al., 2018). We then sought to present a set of cross-disorder findings that were both analytically defensible and most directly comparable to the gold standard univariate GWAS currently available. We plan to offer stronger justification for the inclusion of self-report cohorts, as outlined here, in an updated Method section for resubmission.*

**3. Qtrait heterogeneity index: I believe this is the same index as published in the original genomic SEM paper (Q SNP)? Am I right to understand that in fitting the model where the external trait predicted the individual disorders of a given factor, its effect on the factor and the factor's effect on individual disorders are already fixed (named Step 1 in original genomic SEM paper)? If so, please write explicitly in main text and in Figure 2 and Figure S5, also indicating with arrows (perhaps of a different colour) in Figure 2 and Figure S5 which effects are fixed. At the moment, this seems not to be the case, as shown in Figure 2, where I don't see a line going from trait/SNP to factor in the "Independent Pathways" panel, indicating this effect is fixed and modelled when obtaining Chi2 for the independent effects of trait. If this effect is not modelled, then are we comparing effect of trait/SNP on disease vs effect of trait/SNP on factor? If so, a lack of difference in Chi2 between them cannot indicate anything about the relationship between disease and factor with respect to the trait/SNP (trait/SNP can have the same but independent effect on both). A difference in Chi2 can indicate the trait/SNP has different effects on factor and disease, but can't indicate anything about whether the effects on disease are mediated through the factor. Can the authors clarify? It would be great if the authors can explicitly write out how the Qtrait p values are calculated, as that would make all these confusion go away.**

*The reviewer is correct to point out that $Q_{Trait}$ is akin to the $Q_{SNP}$ metric introduced in the original Genomic SEM paper (Grotzinger et al., 2019). We will update the manuscript based on this comment to clarify that the Q metrics, and corresponding p-values, are calculated via a chi-square difference test for a common pathways model in which the trait/SNP predicts only the factor and an independent pathways model in which the trait/SNP directly predicts the indicators. This is statistically equivalent to the two-step procedure the reviewer references of fixing estimates from the common pathways model and estimating the independent pathways. We will update the Method and Results section text to make these points more clear. This will include providing a reference to McArdle and Goldsmith (1990) that demonstrate common and independent pathways models are nested and, therefore, appropriate for comparison via chi-square difference tests.*

*The reviewer also points out that Q will not be significant in instances when the effect of a trait/SNP has similar effects across the disease traits that define the factor. Mathematically, the single effect of a trait/SNP on a factor is indistinguishable from similar "independent" effects of the trait/SNP on the individual disorders that define the factor. Q is therefore most appropriately viewed similar to most statistical hypothesis tests: as a means of rejecting a null hypothesis (in this case, that the trait/SNP acts on the factor) but not a means of directly confirming the null hypothesis. We will make this interpretational point explicit in our revision.*

**4. Qtrait: "Excluding significant Qtrait correlations (i.e., correlations not operating through the factor), and using the same Bonferroni correction, 17 correlations were significant for Compulsive..." what correlation is this? Geneti. c correlation? Authors please edit text to make things more explicit and understandable.**

*We will carefully go through the text to clarify these instances, including updating the specific phrase highlighted by the reviewer to note that these are genetic correlations*

**5. Figure 3: looking at the personality section, neuroticism is significantly heterogeneous with respect to its correlation with diseases and factor, such that there is a significant portion of its correlaiton with the diseases that is not mediated through factor - is my understanding correct? If so, which disease is this driven by? Is this possible to tell given the way Qtrait is derived? I can't tell as there's nothing about how exactly the estimation of this Qtrait is done in the methods section, please clarify. And how do we understand this result, if this is saying that there is a significant part of neuroticism-internalizing disorder correlations that cannot be explained by the internalzing factor, though the genetic correlation between neuroticism and internalizing factor is > 0.8 and from Supp Figure 7e it seems both internalising traits have similar genetic correlations with neuroticism - is that right? In my mind this is the most interesting result for internalizing disorders, but the authors make no mention in the main text or discussions. Exactly the same situation for agreeableness with neurodevelopmental disorders (though negative effect). The authors even wrote something for an opposite example in the main text "Educational attainment (EA) evinced a particular pattern of genetic associations with the individual compulsive disorders that were inconsistent with their operation via the Compulsive disorders factor, where AN was more positively associated relative to OCD and TS". I find it necessary to discuss at least these two salient observations I pointed out. It is a pity not to talk about any of this in the discussion section.**

*In line with Reviewer 1's third comment, we plan to update the Method section to provide greater clarity on the $Q_{Trait}$/$Q_{SNP}$ metrics. We will clarify that a large correlation estimated between an external trait and factor is not an indication that $Q_{Trait}$ would be expected to be nonsignificant as the reviewer suggests. Indeed, we introduce $Q_{Trait}$ to guard against identifying large genetic correlations between external traits and factor that are ultimately only driven by a subset of the disease traits that load on the factor. As part of the planned revision, we will also update Supplementary Figure 7 to depict the genetic correlation point estimates with error bars. These figures currently plot the genetic correlation z-statistics, which may cause confusion for interpreting $Q_{Trait}$ findings. As requested, we will also provide greater discussion in the Results and Discussion sections to highlight the specific examples of agreeableness and neuroticism.*

**6. Accelerometer data: I am not sure what we've learnt from this. The authors also don't discuss this at all later. If the authors have insights, please add to discussions, otherwise I am unsure why this analysis is in the paper.**

*Dysregulation in daily patterns of movements have been posited to reflect a general risk pathway for psychiatric disease. As the accelerometer results show both divergent patterns of findings across the factors, and convergent patterns for the disorders within a factor, this provides substantial evidence for the validity of the factor model. We will be sure to clarify this point in a revised Discussion section.*

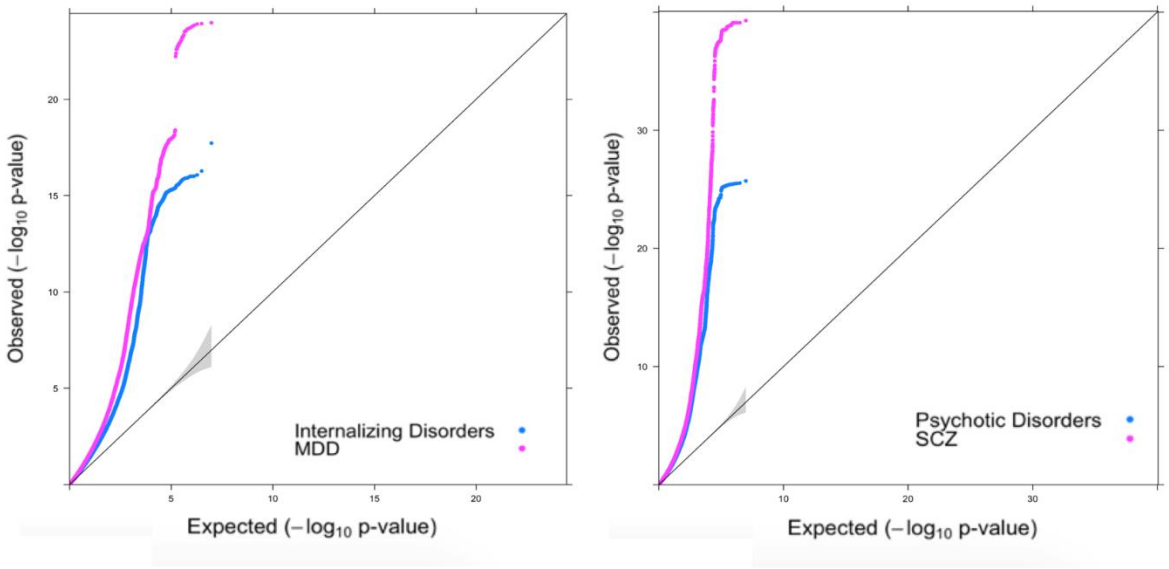
**7. Stratified Genomic SEM: the authors wrote that "Stratified Genomic SEM models that allowed variances of the common genetic factors, and disorder-specific effects, to vary across annotations to examine whether the degree of risk sharing and differentiation is enriched across disorders". Does this mean the method is able to identify the degrees of enrichment in variance attributable to common genetic factors, and disorder specific effects, respectively? From Figure 5 it seems the authors have obtained the enrichment for annotations for the factors identified through previous analyses rather than the individual diseases - taking this as the enrichment results for common genetic factors, where are the results for the disease specific effects? Are those estimated? If not, I can't tell what the difference of this analysis is from S-LDSC on the factors, and I would find the description misleading. Please clarify. Further, in terms of the results, as the effective sample size for each factor used in the enrichment analysis are different, the power to identify significant enrichments for each factor would be different. Given the authors are using summary statistics for all analyses and can't down-sample, please discuss how difference in power affect results of this analysis.**

*We are grateful to the reviewer for raising this particular point. For the resubmission we will also examine enrichment of the residual variances of the different psychiatric traits. This will allow us to determine whether certain annotations are enriched for genetic signal shared across the traits, signal unique to a given trait, or both. We will also note as an interpretive limitation the power differences across the factors.*

**8. Multivariate GWAS: QQplots for all GWAS show significant inflation, and though the authors have found that the intercept from LDSC were close to 1. However, an intercept close to 1 only suggest that the inflation wasn't likely caused by population stratification, and cannot account for**

**inflation due to all statistical errors. Mostly when inflation was due only to polygenicity and not other errors, the qqplots would be inflated at the low P values (high log P), but the qqplots shown in Supplementary figure 23 all deviate from the null even at the high P values (low log P), making me very suspicious the tests are not well calibrated. The authors plot two qqplots in Supplementary figure 23, corresponding to figure 6 in the main text - one for the factor GWAS (blue), and the other for Qtrait tests (magenta). For the Qtrait P values, how are they calculated (I already asked above), are they well calibrated, and how did the authors check for calibration of these p values? I looked into the original genomicSEM paper in addition to this paper, and didn't see any tests for calibration of Qtrait. Similarly, can the authors show that the factors capture the shared effects of triats they load onto, and do not inflate their effects? It would be hard to trust these values without demonstration they are calibrated.**

*With respect to the concern raised about inflation observed in the QQ-plots, the QQ-plots directly below show similar inflation for both univariate GWAS for traits that load on the factor and the psychiatric factors themselves. This strongly suggests that the concern raised by Reviewer 1 is not unique to Genomic SEM, but rather is a reflection of the highly powered nature of the included GWAS statistics on which Genomic SEM is based. The data have been carefully QCd for a variety of confounds (including technical confounds such as batch effects) for the original univariate studies from which we draw and integrate summary data, with the overall consensus being that the early lift-off of QQ plots is largely attributable to the polygenic nature of psychiatric traits. We will include the univariate psychiatric traits in the Figure S23 QQ-plots, as below, in combination with explication of this issue so that this point is clear.*




*The reviewer also notes concerns about how we can be certain that the factors capture shared effects of the factor indicators. In the original Genomic SEM paper (Grotzinger et al., 2019), we demonstrate that the method appropriately accounts for sample overlap, that the standard errors of model estimates and of $Q_{SNP}$ are well calibrated, and that Genomic SEM accurately captures the population generating model. In*

*addition, Grotzinger et al. (2019) report that in a real-data analysis a polygenic score derived from Genomic SEM summary statistics better predicts the individual traits that define the factor than PGS constructed from the summary statistics for the individual traits. This strongly indicates that the power gains obtained via multivariate analysis are not illusory. We also note that the $Q_{SNP}$ statistic is developed to identify and prune out SNPs that are highly disorder specific to guard against inflating the shared signal. We plan to include these details in the Method section so that prior published work done to validate Genomic SEM is clear to readers.*

*The reviewer also requests that we provide demonstration of the calibration of the Q statistic. This was originally presented as a test of heterogeneity of SNP effects, and we will revise the current manuscript to make clear that the mathematical properties of $Q_{Trait}$ are identical to those of $Q_{SNP}$. In the original 2019 paper we provide simulations showing that the null of Q is chi-square distributed, indicating calibration of the test statistic itself. As requested in other comments from the reviewer, we will update the Method section to provide greater detail on how Q is calculated and we will report simulations to demonstrate how Q performs under a variety of population generating models.*

**Minor**

**1. Figure S8 has nano-sized fonts, really difficult to read, please consider breaking up into multiple pages like Figures S4, S7, S9 etc**

*We will increase the font size and break up this figure across multiple panels.*

**2. I am not opposed to polar plots but think they are unnecessarily hard to read. Normal bar plots would do the jobs just as nicely.**

*As requested, we will update our plots to be more accessible.*

**3. For most of the supplemental figures, perhaps it's much easier to represent these as tables.**

*We note that many of the figures in the Online Supplement (Miami plots, enrichment findings, simulation results) have corresponding supplementary tables. That said, we plan to include corresponding supplementary tables for all relevant supplemental figures.*

**Reviewer #2:**
**Remarks to the Author:**

**The authors use genomic structural equation modeling to infer 4 factors underlying 11 psychiatric disorders and draw inferences about the genetic architecture of these 4 factors. They then conduct a GWAS of the 4 factors, identifying 152 loci including 20 novel loci.**

**Section identifying the 4 factors:**

**(1) This is only a mild extension of the results of Lee et al. 2019 Cell (ref. 13), who analyzed 8 of these psychiatric disorders and identified 3 of these factors. The authors are transparent about this.**

**Thus, this section should be made much shorter. Confirmatory factor analyses and other subtle details can be moved entirely to the supplement. The comments about Cai et al. 2020 Nat Genet (ref. 28), though appropriate, can be moved to the Discussion.**

*We will move technical details about the confirmatory factor analysis to the Online Supplement as the reviewer suggests. Please note that by expanding the range of disorders, we have elaborated the original 3 factor model reported in Lee et al. (2019) to now include an important fourth factor: internalizing disorders. This factor, which includes major depressive disorder (which had previously clustered with the psychotic disorders), represents an important dimension of variation of major clinical significance. Please also note that the original PGC Cross-disorder paper (CDG1, 2013) consisted of five major disorders. The major update by Lee et al. (2019) for the second major PGC Cross-disorder effort (CDG2) then included three additional disorders, for a total of eight disorders. Similarly, we add an additional three disorders for a total of 11 disorders, and substantially update many of the sample sizes relative to those from the Lee et al. (2019) publication*

**(2) The p-factor model seems to be of low interest. It has greater complexity, and has close to zero genetic associations in the GWAS part of the paper. Either all content on the p-factor model should be moved to the supplement, or considerable additional justification is needed as to why it is important in this paper.**

*We agree with the reviewer that our results with respect to the p-factor (e.g., close to zero GWAS genetics associations, no meaningful mediation of the relationship between external traits and psychopathology as measured by $Q_{Trait}$) indicate that the p-factor is of little theoretical or pragmatic utility. If these results were motivated by a novel hypothesis on our part, we would agree that the fact that they were unsupported would limit their interest. However, the p-factor is a construct that is widely cited within the literature to the extent that it is often implicitly treated as "real" or at least meaningful. For example, the 2014 Caspi et al. paper (DOI: 10.1177/2167702613497473) naming the p factor has been cited over 1,400 times according to Google scholar. Thus, we believe that our lack of support for the p-factor in the context of the very high general interest this transdiagnostic construct, is strong rationale for including the results in the main text.*

**(3) Minor comments: The authors should include a main Table that includes all of the information in Table 1 of Lee et al. 2019 Cell (ref. 13). Also, Figure 1C and Figure 1D should be moved to the supplement. (Not clear what Figure 1D refers to, as the text does not mention a "bifactor model".) The text fragment "prior literature indicating a high-order transdiagnostic 'p-factor'" requires a reference, perhaps ref. 6 or ref. 6-8 as cited previously.**

*We will move Table S1, which includes the sample size and contributing study information for each disorder, to the main text and add population prevalences and estimated heritabilities to the table to parallel Table 1 from Lee et al. (2019). We will also include a reference for the noted text fragment. In line with our response to Reviewer 2 point 2, our aim would be to retain Figure 1C and 1D while providing further justification for including the p-factor analyses in the main text.*

**Section on genetic correlation of the 4 factors with external traits:**

**(4) Qtrait is not a compelling metric, both because statistical significance is sample size dependent, and because the null hypothesis that the genetic correlation between an external trait and the 11 psychiatric disorders operates entirely through one of the factors is not a plausible null hypothesis, such that its violation is not meaningful. It might be more interesting to assess what proportion of the genetic correlation between an external trait and the 11 psychiatric disorders operates through one of the factors (or all 4 factors). Also, Figure 2 should be moved to the supplement.**

*$Q_{Trait}$ quantifies the extent to which the pattern of genetic correlations with the individual disorders is proportional to the factor loadings themselves. This allows us to directly test the extent to which the trait's association with the disorders can be plausibly attributed to a single effect on the factor, which is different from examining whether the genetic correlations are exactly equal across all 11 psychiatric disorders. We agree that the latter would be a somewhat uninformative null hypothesis, while the former helps to clarify what associations with external correlates are likely to operate through overarching factors of psychiatric risk. We will make the null of $Q_{Trait}$ more explicit in the main text. Moreover, we will re-iterate the explication from the supplement to de la Fuente et al. (2020), that makes clear that because Q is a hypothesis test regarding the extent to which regression coefficients (and not Z statistics) conform to a factor loading pattern, Q is not biased by differential power across traits.*

**(5) Some of the genetic correlations of the 4 factors with external traits are interesting (particularly for the accelerometer traits), but it is unclear whether the insights gained are more impactful than the insights gained by assessing (only) the genetic correlations of the 11 psychiatric diseases with external traits (for example, in Figure 4, results for 4 factors look similar to results for corresponding psychiatric diseases). It would be important to discuss this.**

*The reviewer's observation that many of the genetic correlations with the 11 psychiatric diseases appear similar to the correlations identified for the 4 factors speaks to the utility of using Genomic SEM to understand these patterns of correlations. We will clarify the value of testing validity of the factors using the evidence from external patterns of genetic correlation. This is particularly compelling because the factor model itself was developed on the basis of internal patterns of genetic correlations among the disorders. That is, given similar patterns observed in the genetic correlation matrices, Genomic SEM allows us to formally quantify the genetic correlation between psychiatric traits and external indicators and to test using $Q_{Trait}$ whether these patterns are significantly different across the psychiatric traits. We will include this justification in the main text.*

**Section on genetic enrichment of the 4 factors in annotations:**

**(6) "Stratified Genomic SEM allows us to ask whether pleiotropic loci are enriched within particular annotations": I agree with this comment. However, why not just stratify genetic covariance by annotation, which also assesses whether pleiotropic loci are enriched within particular annotations? (The authors actually develop this approach by developing stratified multivariate S-LDSC as the first step of their Stratified Genomic SEM method, although this is not made clear until the Methods section. Also, Lu et al. 2017 Am J Hum Genet is a published method for stratifying genetic covariance by annotation that should be cited and discussed.) It is unclear whether the insights gained from the factor approach are more impactful than the insights gained by (only) stratifying genetic covariance by annotation. It would be important to discuss this.**

**Specifically, are there any annotations for which the genetic enrichment for one of the 4 factors is significantly larger in magnitude than the genetic enrichment for any of the 11 psychiatric disorders?**

*We plan to make more explicit the validation of stratified multivariate S-LDSC as part of the development of Stratified Genomic SEM, and include the noted reference, in our resubmission. There are a number of advantages to examining enrichment at the level of the factors, rather than at the level of individual variables or pairwise combinations of variables, that we will discuss in our revision. Enrichment of genetic covariance provides information about the gene categories in which pleiotropic signal, as opposed to heritability itself, is disproportionately localized. However, when data become multivariate, bivariate pairs of variables rapidly expand, which renders examination of enrichment itself high dimensional. Genomic SEM resolves this issue by allowing for enrichment to be examined at general dimensions of genetic variation spanning large constellations of (rather than simply pairs of) phenotypes. In response to comments from Reviewer 1, we will also now report enrichment of the residual variance of the indicators. In the context of Stratified Genomic SEM, this allows us to ask whether there is enrichment of signal that is unique to one of the indicators, which would not be possible by performing standard S-LDSC on the factors or examining stratified genetic covariance in isolation.*

**(7) Minor comment: the term "genetic correlation" usually refers to a relationship that has consistent sign across the genome. The term "pleiotropic" usually refers to a relationship that may have varying sign across the genome. This paper seems to focus on relationships that have consistent sign across the genome. Thus, either the term "genetic correlation" (or covariance) should be consistently used, or at least the terminology (including the use of the term "pleiotropic") should be carefully defined/explained.**

*We will update the text to be more clear when we are referring specifically to the case of unidirectional pleiotropy.*

**Unstructured multivariate GWAS and structured multivariate GWAS sections:**

**(8) The "unstructured" GWAS (which does not make use of the 4 factors) identified 184 loci including 39 novel loci. The "structured" GWAS (which does make use of the 4 factors) identified 152 loci including 20 novel loci. Why is the structured GWAS preferred? Are its results more actionable, and if so why?**

*In line with prior cross-disorder efforts, the unstructured GWAS allows us to ask what SNPs have an effect on any of the psychiatric traits regardless of direction. This allows us to produce an exhaustive list of SNPs relevant to psychiatric risk. In comparison, the structured GWAS adds in the nuance of identifying SNPs that are specifically relevant to each the four factors. We do not view one as more actionable than the other, but rather as providing different levels of information, each of which is useful in its own right. We plan to include in an updated revision that provides greater clarification on how we suggest interpreting these two sets of results in the main text.*

**(9) How do the unstructured GWAS method and the structured GWAS method compare to the considerable literature on multi-trait GWAS methods that has been published, including**

**Nieuwboer et al. 2016 Am J Hum Genet (not cited, same last author as this paper) and Turley et al. 2018 Nat Genet (not cited) and Lee et al. 2019 Cell (ref. 13) and Grotzinger et al. 2019 Nat Hum Behav (ref. 44; same first author as this paper)? In particular, is the structured GWAS method proposed in this paper identical to the structured GWAS method proposed by ref. 44? Although ref. 44 is cited in the structured GWAS method, this is not explicitly clear.**

*The reviewer is correct that the structured GWAS is estimated using Genomic SEM, which we introduced in our 2019 paper in Nat Hum Behavior. We will be sure to make this point more explicit in the main text. We are happy to cite the work of Nieuwboer et al. 2016, but this work is not actively being developed to be used for structure multivariate analyses, and we have previously shown in Grotzinger et al. (2019), many of its relevant capabilities are subsumed by the Genomic SEM framework. Please see response to the next question for further detail on how we will compare our methodology to other multi-trait methods.*

**(10) For these sections to have high methodological impact, it would be critically important to add simulations comparing the calibration and power of the proposed structured GWAS method to other multi-trait methods in simulated data, and explain (with justification) in which scenarios the method provides an increase in power.**

*As one of the major updates to a resubmission, we will include simulations to benchmark the performance of Genomic SEM under different population generating conditions. We will specifically compare structured and unstructured Genomic SEM to the algorithms for two main, multi-trait methods: MTAG (Turley et al., 2019) and the multi-trait version of METAL (Willer et al., 2010). We will compare Genomic SEM to MTAG as this specifically represents a method that, like Genomic SEM, takes into account the genetic covariance across traits estimated by LDSC. We chose METAL as an additional multi-trait comparison point as this is a commonly used software that is capable of taking into account sample overlap. As a major focus in the field is to determine the generality vs. specificity of genetic associations, our goal will not be to produce a simple head-to-head comparison of power, but to determine the conditions under which pleiotropic and non-pleiotropic variants will be correctly and incorrectly identified by each method.*

**(11) For these sections to have high biological impact, it would be critically important to add considerable biological interpretation of the novel findings. Also, it would be preferred to empirically replicate novel findings in independent data sets, as an additional validation.**

*We will substantially update the Discussion to consider the prior literature relevant to the pattern of enrichment findings identified in the current analyses. While we would ideally replicate these findings as the reviewer suggests, many of the included disorders represent the only GWAS of sufficient sample size to be included in analyses like Stratified Genomic SEM or LDSC more generally. We will be sure to note this as a limitation and avenue for future analyses in any resubmission.*

**(12) QSNP is not a compelling metric, because the null hypothesis that the association between a SNP and a psychiatric disorder operates entirely through one of the factors is not a plausible null hypothesis, such that its violation is not meaningful. It might be more interesting to assess what proportion of the association between a SNP and a psychiatric disorder operates through one of the factors (or all 4 factors).**

*Please see our response to Reviewer 2 comment 4 that offers the same critique of the $Q_{Trait}$ metric.*

**(13) Minor comments: definitely state in the Results section that 4,775,763 SNPs were tested for association (intersection across 11 psychiatric disorders) and a genome-wide significance threshold of 5e-08 was used. It would be good to justify the significance threshold by citing previous references in which this rather lax threshold was used. Also, "genome-wide S-LDSC matrix": does this refer to using a genetic covariance matrix derived from S-LDSC instead of LDSC? This should be clarified.**

*We will update the text in the Results section to state that we include the 4,775,763 SNPs, and clarify that genome-wide S-LDSC matrix does refer to the genetic covariance matrix derived from the S-LDSC annotation that includes all SNPs. We will also update the text to provide the original citation that justifies using 5e-8 as the significance threshold for common variant analyses in European populations (Pe'er et al., 2007), and note that even the most recent PGC GWAS efforts use this same 5e-8 threshold (e.g., Ripke et al., 2020).*

**The last two Results sections add little to the paper, and should be moved to the supplement.**

*In the interest of including more details on the Stratified Genomic SEM simulations, and additional simulations validating the use of Genomic SEM for multivariate GWAS, we are happy to move these two sections to the Online Supplement.*

## References

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, *381*(9875), 1371-1379.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., ... & Koellinger, P. D. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour*, *3*(5), 513-525.

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., ... & Alloza, C. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*, *22*(3), 343-352.

Lee, P. H., Anttila, V., Won, H., Feng, Y. C. A., Rosenthal, J., Zhu, Z., ... & Wang, M. M. J. (2019). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*, *179*(7), 1469-1482.

McArdle, J. J., & Goldsmith, H. H. (1990). Alternative common factor models for multivariate biometric analyses. *Behavior Genetics*, *20*(5), 569-608.

Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. (2007). Estimation of the Multiple Testing Burden for Genomewide Association Studies of Common Variants. *Nature Precedings*, 1-1.

Ripke, S., Walters, J. T., O'Donovan, M. C., & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*.

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Adams, M. J., ... & Deary, I. J. (2019). Genome-wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. *American Journal of Psychiatry*, *176*(2), 107-118.

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... & Magnusson, P. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics*, *50*(2), 229-237.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190-2191.

**Decision Letter, first revision:**

IMPORTANT: Please note the reference number: NG-A55864R-Z Grotzinger. This number must be quoted whenever you communicate with us regarding this paper.

24th December 2020

Dear Andrew,

Thank you for asking us to reconsider our decision on your manuscript "Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis". I have discussed your responses and proposed revisions with my editorial colleagues, and we would like to invite you to revise your manuscript along the lines that you propose for further editorial consideration and peer review.

When preparing a revision, please ensure that it fully complies with our editorial requirements for format and style; details can be found in the Guide to Authors on our website (http://www.nature.com/ng/).

Please be sure that your manuscript is accompanied by a separate letter detailing the changes you have made and a point-by-point response to the referee comments. At this stage we will need you to upload:

1) A copy of the manuscript in MS Word .docx format.

2) The Editorial Policy Checklist:
https://www.nature.com/documents/nr-editorial-policy-checklist.pdf

3) The Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf

(Here you can read about the role of the Reporting Summary in reproducible science:
https://www.nature.com/news/announcement-towards-greater-reproducibility-for-life-sciences-research-in-nature-1.22062 )

Please use the link below to be taken directly to the site and view and revise your manuscript:

[REDACTED]

With best wishes,
Kyle

Kyle Vogan, PhD
Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665

**Author Rebuttal, first revision:**

**Reviewer #1:**
**Remarks to the Author:**
**Comments**

**This paper performs genomicSEM on 11 psychiatric traits and identified four broad factors (Neurodevelopmental, Compulsive, Psychotic, and Internalizing) that model their genetic correlations. They further checked for genetic correlation of each factor with a variety of traits, and functional enrichment of genes expressed in different tissue types for each factor. Most importantly, they looked into the genetic factors who contributions to each disorder cannot be accounted for by the factors that load onto them, and performed MR analysis to show that at least at one such locus, there is evidence of causality where its effects on alcohol abuse may account for its effects on internalizing and neurodevelopmental disorders. Given the genetic effects that cannot be attributed to the factors, and that are heterogeneous among the diseases each factor load onto, the authors have concluded there is little utility in one common factor for psychiatric disorders. My concerns lie in the use of self-report data, and whether their results are well calibrated. I have written my concerns below for the authors' considerations.**

**Major**

**1. As the authors mentioned, misdiagnoses and other non-specificity incurred through using self-reported phenotypes may affect results of the analysis. Though the authors have performed the confirmatory factor analysis removing the with self-reported cohorts, and obtained similar results (with larger difference for factor loadings on ANX, lowest congruence coefficient for internalizing factor), I would rather like to see exploratory factor analysis done without the self-report cohorts - would the same factor models be obtained, and which would be the best fit in confirmatory factor analysis?**

*The reviewer's point is well taken and we have provided a much more comprehensive survey on the effects of including self-report cohorts in the updated manuscript. We have re-run the EFA analyses excluding self-report cohorts and report the full set of results in the Online Supplement. We find that a highly similar correlated factors solution is obtained when using the restricted dataset. In addition, we find that when fitting both this factor solution and the correlated factors solution identified using the full dataset to the restricted data that the solution identified using all data provides slightly better model fit. Finally, we examine the 152 hits identified in the full dataset in the dataset excluding self-report cohorts and identify highly concordant signal in the restricted dataset. This suggests that results are not qualitatively shifted by including these self-report cohorts. We write specifically in the Online Supplement (page 5):*

*"Genomic SEM Estimates Excluding Self-Report GWAS. In a sensitivity analysis, we re-examined the Genomic SEM factor solutions when excluding GWAS summary statistics that included cohorts for which the psychiatric phenotypes were based primarily on self-report items not directly assessed by a clinician. This involved excluding the UK Biobank samples from MDD, ANX, and ALCH, and the 23andMe cohorts from MDD and ADHD. This reflected an 81% reduction in effective sample size for MDD, an 82% reduction for ANX, a 24% reduction for ADHD, and an 84% reduction for ALCH. To begin, we examined the heatmap of genetic correlations among the 11 traits, along with the difference in genetic correlations relative to genetic correlations estimated using all cohorts. We observe similar patterns of clustering among the traits (Supplementary Figure 1). Relative to using all cohorts, these analyses produced slightly larger genetic correlations for MDD and slightly smaller correlations for ANX.*

*We conducted a new EFA excluding the self-report cohorts using the same procedure of fitting the EFA in odd chromosomes and the CFA in even chromosomes. These analyses revealed a correlated factors model with four factors to be the best fitting model, with this model fitting the data well in both even chromosomes, $\chi^2[35] = 135.70$, AIC = 197.70, CFI = .907, SRMR = .104; and all chromosomes, $\chi^2[35] = 209.54$, AIC = 271.54, CFI = .936, SRMR = .078 (Supplementary Figure 1c; Supplementary Table 51). This factor structure was highly similar to that identified using all cohorts, with the factors again best characterized as reflecting compulsive, psychotic, neurodevelopmental and internalizing disorders. The one notable exception was cross-loadings of both MDD and ANX on the Compulsive disorders factor. A hierarchical model fit overtop this factor structure fit the data relatively worse in both even chromosomes, $\chi^2[37] = 170.30$, AIC = 228.30, CFI = .878, SRMR = .147; and all chromosomes, $\chi^2[37] = 231.59$, AIC = 289.59, CFI = .929, SRMR = .103.*

*We went on to estimate the parameters from the final confirmatory correlated factor model represented in Figure 1 using this more restricted dataset. Overall, both factor loadings and factor correlations from this restricted dataset were highly similar to those for the full dataset, and fit the data well, $\chi^2[33] = 189.48$, AIC = 255.48, CFI = .942, SRMR = .098; albeit with a lower loading of ANX on the Internalizing disorders factor (Supplementary Figure 1). We additionally used the restricted dataset to estimate a five-factor orthogonal EFA model, which was the model that served as the basis for the final confirmatory factor model in the main set of analyses. To quantify the similarity of EFA solutions across the full and restricted datasets, we computed factor congruence coefficients using the R psych package. Congruence coefficients index the similarity between factor solutions, with possible values ranging between -1.0 and +1.0. A congruence coefficient greater than .90 indicates an extremely high level of similarity of the factors, and values above .84 are considered reasonably similar. The congruence coefficients were .92 for the Compulsive disorders factor, 1.0 for the Psychotic disorders factor, .93 for the Neurodevelopmental disorders factor, and .85 for the Internalizing disorders factor. Of note, the factor solution identified using all cohorts provided better fit to the data excluding self-report cohorts than the factor solution identified using an EFA in self-report cohorts only reported above. In order to provide a more direct comparison to results using the full dataset, and owing to the better model fit, the correlated factor model identified using the full dataset was carried forward to examine GWAS hits in the restricted dataset.*

*As a final set of sensitivity analyses, we reexamined the SNP effects for the 154 hits identified from the correlated factors model estimated in the restricted dataset. All hits for the Compulsive and Psychotic disorders factor were also identified as hits using the restricted dataset, 8 out of 9 hits were genome-wide significant for the Neurodevelopmental disorders factor, and none of the 44 hits were estimated as genome-wide significant for the Internalizing disorders factor (Supplementary Table 3). However, plotting the distribution of effects indicate clear signal for these 44 loci in the restricted dataset relative to the estimated SNP effects for a random subset of 500 SNPs. Moreover, there was extremely high concordance for this subset of SNPs for the estimated factor betas across the full and restricted datasets ($r \geq .94$ Supplementary Figure 2). In addition, the hits identified in the full dataset were neither $Q_{SNP}$ hits nor characterized by robust $Q_{SNP}$ signal in the restricted dataset (Supplementary Table 3). Finally, we note that there were only 2 independent loci for MDD and 1 independent locus for ANX in the restricted sample for the listwise deleted set of SNPs present across the 11 psychiatric disorders. The absence of Internalizing factor hits in the restricted dataset, therefore, appears to largely reflect an attenuated signal as a result of a substantial reduction in sample size."*

*Collectively, the sensitivity analyses excluding self-report cohorts reported above indicate that results do not qualitatively shift by including self-report GWAS estimates. We also believe that the decision to include certain self-report cohorts in the main analysis is defensible for a number of reasons. We offer stronger justification for the inclusion of self-report cohorts in an updated Method section (page 34):*

*"The current analyses included GWAS summary statistics produced using self-report items not directly assessed by a clinician for MDD, ANX, ALCH, and ADHD. The inclusion of these cohorts was based on the large genetic correlations between the clinically diagnosed and self-report GWAS, the increased mean chi-square when meta-analyzing self-report and clinical diagnosis GWAS (Supplementary Table 45), and a general trend in psychiatric genomics to include self-report cohorts in the primary GWAS studies being published. In some cases, multiple self-report options were available, in which case*

*phenotypes were chosen based on the field standard and prior findings. For example, the choice to use the broad depression phenotype from UK Biobank and self-report 23andMe MDD phenotype was based on the inclusion of both phenotypes in the most recent GWAS of MDD[75] and of the latter phenotype in the prior PGC GWAS of MDD.[22] In addition, Wray et al.[22] find that polygenic scores constructed from MDD 23andMe summary statistics predict equal, or greater, amounts of out-of-sample variance in MDD phenotypes than PGS constructed from PGC case/control summary statistics. Moreover, they find that the meta-analyzed summary statistics across both 23andMe and PGC cohorts predicted the greatest amount of variance. We note also that the meta-analysis between PGC Alcohol Use Disorder and UKB self-reported alcohol use is limited to self-reported problematic alcohol use (as assessed by the AUDIT-P) and not alcohol consumption (as assessed by the AUDIT-C). This is based on prior work indicating stronger genetic correlations between self-reported problematic alcohol use and alcohol dependence relative to self-reported alcohol consumption.[76]"*

**2. How does difference in sample size between the different diseases affect the analysis - they would lead to different standard errors in genetic covariance estimates, and as explained in the original genomic SEM paper, the WLS produce a solution that is dominated by the patterns of association involving the most well-powered GWAS... maximum likelihood estimates will be most pronouced when there is lower sample overlap and the contributing univariate GWAS differ substantially in sample size. Related to above comment, removing self-reported cohorts would reduce sample size of MDD, ANX, and ALCH significantly (>80%), and less so for ADHD (>20%) - the sample size for these diseases are larger than those of other diseases before removing self-report cohorts. Would reducing their sample size to a more comparable level as other diseases help the analysis? Given this, I have further reason to believe the exploratory step should be done again using only the cohorts without self-report.**

*As noted above, we have provided a comprehensive set of sensitivity analyses, including conducting the exploratory step excluding self-report cohorts, and additionally provide stronger justification for including the self-report cohorts. We thank the reviewer for pointing out the interpretive consideration around sample size differences and have updated the Method section to provide an explanation of how to interpret WLS estimates in Genomic SEM in the context of differing sample sizes, as per below. It is also of note that our sensitivity analyses in which self-report cohorts were excluded resulted in substantial differences in relative sample sizes across traits, but continued to produce a very similar factor solution, thereby demonstrating that the identified factor structure is not driven by differences in sample size across disorders. We now write on page 35:*

*"All CFAs were fit using the Weighted Least Squares (WLS) estimator in the GenomicSEM R package described above, which uses the inverse of the diagonal of the sampling covariance (V) matrix to weight the discrepancy function. This works to prioritize reducing model misfit for those cells in the genetic covariance matrix that are estimated with greater precision, with the desirable result of generally decreasing the sampling variance of parameter estimates in Genomic SEM. It should be noted that WLS estimation does not necessarily produce a solution whereby the better powered GWAS have larger factor loadings. In instances where traits with better-powered GWAS estimates evince lower genetic correlations with other included traits, WLS estimation will produce a solution that prioritizes lower factor loadings for these traits and consequently minimize their downstream influence on multivariate GWAS estimates."*

**3. Qtrait heterogeneity index: I believe this is the same index as published in the original genomic SEM paper (Q SNP)? Am I right to understand that in fitting the model where the external trait predicted the individual disorders of a given factor, its effect on the factor and the factor's effect on individual disorders are already fixed (named Step 1 in original genomic SEM paper)? If so, please write explicitly in main text and in Figure 2 and Figure S5, also indicating with arrows (perhaps of a different colour) in Figure 2 and Figure S5 which effects are fixed. At the moment, this seems not to be the case, as shown in Figure 2, where I don't see a line going from trait/SNP to factor in the "Independent Pathways" panel, indicating this effect is fixed and modelled when obtaining Chi2 for the independent effects of trait. If this effect is not modelled, then are we comparing effect of trait/SNP on disease vs effect of trait/SNP on factor? If so, a lack of difference in Chi2 between them cannot indicate anything about the relationship between disease and factor with respect to the trait/SNP (trait/SNP can have the same but independent effect on both). A difference in Chi2 can indicate the trait/SNP has different effects on factor and disease, but can't indicate anything about whether the effects on disease are mediated through the factor. Can the authors clarify? It would be great if the authors can explicitly write out how the Qtrait p values are calculated, as that would make all these confusion go away.**

*The reviewer is correct to point out that $Q_{Trait}$ is akin to the $Q_{SNP}$ metric introduced in the original Genomic SEM paper (Grotzinger et al., 2019). We have updated the manuscript based on this comment to clarify that the Q metrics, and corresponding p-values, are calculated via a chi-square difference test for a common pathways model in which the trait/SNP predicts only the factor and an independent pathways model in which the trait/SNP directly predicts the indicators. This is mathematically equivalent to the two-step procedure the reviewer references of fixing estimates from the common pathways model and estimating the independent pathways. We have updated both the Results and Method sections to make these points clearer. We specifically write on page 7 of the Results:*

*"To evaluate the extent to which each of the 49 biobehavioral traits operated through the factor, we calculated $\chi^2$ difference tests comparing a model in which the trait predicted the factor only, to one in which it predicted the individual disorders of a given factor (or, the first-order factors, in the case of analyses using the p-factor model; Figure 2; Supplementary Figure 5). As this is mathematically equivalent to the $Q_{SNP}$ metric introduced in the original Genomic SEM publication,[14] we term the $\chi^2$ difference across these two models the $Q_{trait}$ heterogeneity index. A significant $Q_{trait}$ index indicates that the pattern of associations between the individual disorders and the external trait is not well-accounted for by the factor. Using a Bonferroni correction, 7/49 correlations were significant for $Q_{trait}$ for the Compulsive factor, 18/49 for the Psychotic factor, 39/49 for the Neurodevelopmental factor, 17/49 for the Internalizing factor, and 38/49 for the p-factor (Supplementary Table 5). Excluding genetic correlations significant for $Q_{trait}$ (i.e., genetic correlations not likely to operate through the factor), and using the same Bonferroni correction, 17 genetic correlations were significant for the Compulsive factor, 12 for the Psychotic factor, 5 for the Neurodevelopmental factor, 20 for the Internalizing factor, and 3 for the p-factor."*

*We have also added the following to the figure note for Figure 2:*

*"Q is estimated here using a $\chi^2$ difference test across the common and independent pathways models; this is statistically equivalent to the 2-step procedure outlined in the original Genomic SEM[14] publication for calculating $Q_{SNP}$."*

*Finally, we write in the Method section (page 36):*

***"Estimation of Q Metrics***

*We compute heterogeneity statistics for both associations with external traits ($Q_{Trait}$) and individual SNPs ($Q_{SNP}$). These index violation of the null hypothesis that a given trait or SNP acts through a given factor. Put another way, it quantifies whether the external trait or SNP is more likely to operate through the common pathways of the psychiatric factors, or the independent pathways of individual disorders. These Q metrics thereby identify instances when associations with a trait or SNP do not plausibly operate on the individual phenotypes exclusively by way of associations with common factor(s), and may be highly specific to the individual disorder. Four separate, follow-up models were estimated in which the SNP or trait predicted three of the overarching factors and the indicators of the remaining fourth factor (see Supplementary Figure 5 for $Q_{Trait}$ path diagrams; Supplementary Figure 45 for $Q_{SNP}$ path diagrams). Computing the nested $\chi^2$ difference test between the common pathways model, in which the SNP or trait predicted all four factors, to one of these four, follow-up, independent pathways models produces a factor-specific Q metric. We note that it has been previously demonstrated that common and independent pathways models are nested and, therefore, appropriate for comparison via the nested $\chi^2$ difference tests[78] used to compute Q metrics here.*

*We calculate model $\chi^2$ for both the common and independent pathways models using the two-step procedure described in Grotzinger et al. (2019).[14] In Step 1 of this procedure a proposed model is estimated. In Step 2, the Step 1 estimates are fixed and the residual covariances and variance of the indicators are freely estimated. The estimates in Step 2 capture both the discrepancy between the model implied and observed covariance matrices, and the corresponding sampling covariance matrix ($V_R$) of R. The $V_R$ matrix has the eigendecomposition:*

$$V_R = (P_1\, P_0) \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} P_1' \\ P_o' \end{pmatrix}$$

*with $P_1$ reflecting a matrix of principal components (eigenvectors) of $V_R$, $E$ a corresponding diagonal matrix consisting of non-zero eigenvalues, and $P_0$ the null space of $V_R$. Projecting $R_i$, the vector of residual covariances estimated in Step 2, onto $P_1$ and adjusting for corresponding eigenvalues produces:*

$$E^{\frac{-1}{2}} P_1' R_i N(0, I_r)$$

*Therefore,*

$$R_i' P_1 E^{-1} P_1' R_i \sim \chi^2(r)$$

*It has been previously confirmed via simulation that this equation produces a $\chi^2$ distributed test statistic.[14] This method of computing $\chi^2$ difference tests across a common pathways model and an independent pathways model to arrive at a Q metric is mathematically equivalent to the procedure outlined for calculating $Q_{SNP}$ in Grotzinger et al. (2019).[14]*

*We note a number of important points to keep in mind with respect to interpreting Q (see de la Fuente et al. [2020][79] for additional explication). First, Q will be most significant for a factor when the vector of observed effects with an external trait or SNP is not proportional to the unstandardized loadings of the disorders on the factor. Consequently, Q is not necessarily significant when the vector of observed*

*external SNP/trait effects is unequal across the disorders as, in many cases, the disorders will also have unequal unstandardized factor loadings. For example, in cases where a particular disorder has a low unstandardized loading relative to the other disorders, we would expect Q to be high for SNPs or external traits that show comparable associations across all disorders. As Q is calculated based on observed beta coefficients, and not z-statistics, this has the desirable property that Q will not increase simply due to differences in power across the univariate GWAS. As an interpretive caveat, we note also that Q will not be significant in instances when the effect of an external trait or SNP has similar, but independent, effects on the disorders that define the factor. In this sense, Q is most appropriately viewed in the same light as many other statistical hypothesis tests: as a means of rejecting the null (i.e., that the trait or SNP acts solely via the factor) but not as a means of directly confirming the null. Indeed, patterns of external associations are generally not expected to conform exactly to the factor model, just as population effects are never expected to be exactly 0. However, by setting stringent significance thresholds we seek to identify via Q those SNPs and external traits that strongly deviate from the factor structure, thereby offering insight into underpinnings of genetic divergence across even highly correlated disorders.*

*For the hierarchical factor structure, we computed the $\chi^2$ difference test for a model in which the SNP or trait predicted only the second-order p-factor, to the model $\chi^2$ for a model in which the SNP or trait predicted only the four, first-order psychiatric factors. For the bifactor model, we compared a model in which the SNP predicted only the p-factor to a model in which the SNP predicted both the p-factor and the remaining four orthogonal factors. For both the hierarchical and bifactor model, Q indexes heterogeneity at the level of the psychiatric factors (i.e., deviation from the null that the SNP or trait operates through the p-factor). Therefore, a significant Q statistic for the hierarchical or bifactor model is likely to identify patterns of external associations that are specific to a subset of the psychiatric factor(s). This is distinct from the interpretation of Q in the context of the correlated factors model, as a significant hierarchical or bifactor Q may still conform to the local structure of one of the correlated factors."*

**4. Qtrait: "Excluding significant Qtrait correlations (i.e., correlations not operating through the factor), and using the same Bonferroni correction, 17 correlations were significant for Compulsive..." what correlation is this? Genetic correlation? Authors please edit text to make things more explicit and understandable.**

*We have carefully gone through the text to clarify these instances, including updating the specific phrase highlighted by the reviewer to note that these are genetic correlations as below (page 7):*

*"Excluding genetic correlations significant for $Q_{trait}$ (i.e., genetic correlations not operating through the factor), and using the same Bonferroni correction, 17 genetic correlations were significant for the Compulsive factor, 12 for the Psychotic factor, 5 for the Neurodevelopmental factor, 20 for the Internalizing factor, and 3 for the p-factor."*

**5. Figure 3: looking at the personality section, neuroticism is significantly heterogeneous with respect to its correlation with diseases and factor, such that there is a significant portion of its correlaiton with the diseases that is not mediated through factor - is my understanding correct? If so, which disease is this driven by? Is this possible to tell given the way Qtrait is derived? I can't tell as there's nothing about how exactly the estimation of this Qtrait is done in the methods section, please clarify. And how do we understand this result, if this is saying that there is a significant part**

**of neuroticism-internalizing disorder correlations that cannot be explained by the internalzing factor, though the genetic correlation between neuroticism and internalizing factor is > 0.8 and from Supp Figure 7e it seems both internalising traits have similar genetic correlations with neuroticism - is that right? In my mind this is the most interesting result for internalizing disorders, but the authors make no mention in the main text or discussions. Exactly the same situation for agreeableness with neurodevelopmental disorders (though negative effect). The authors even wrote something for an opposite example in the main text "Educational attainment (EA) evinced a particular pattern of genetic associations with the individual compulsive disorders that were inconsistent with their operation via the Compulsive disorders factor, where AN was more positively associated relative to OCD and TS". I find it necessary to discuss at least these two salient observations I pointed out. It is a pity not to talk about any of this in the discussion section.**

*In line with Reviewer 1's third comment, we have updated the Method section to provide greater clarity on the $Q_{Trait}$/$Q_{SNP}$ metrics. This includes clarifying that a large correlation estimated between an external trait and factor is not an indication that $Q_{Trait}$ would be expected to be nonsignificant as the reviewer suggests. Indeed, we introduce $Q_{Trait}$ to guard against identifying large genetic correlations between external traits and factor that are ultimately only driven by a subset of the disease traits that load on the factor. We have also updated Supplementary Figure 7 to depict the genetic correlation point estimates with error bars. These figures previously plotted the genetic correlation z-statistics, which may cause confusion for interpreting $Q_{Trait}$ findings. As requested, we also highlight the specific examples of agreeableness and neuroticism. We specifically write in the Results (page 10):*

*"The Neurodevelopmental disorders factor was genetically associated with earlier age at menopause. All other external correlates outside of the psychiatric domain that survived Bonferroni-correction exhibited patterns of associations with the individual neurodevelopmental disorders that were inconsistent with their operation via the factor. Cognitive (e.g., educational attainment, intelligence) and economic outcomes (e.g., own housing outright) had the strongest disorder-specific associations, with positive associations observed for AUT, and negative associations for PTSD and ADHD. In a few instances, PTSD stood apart from the remaining indicators. This included a stronger, negative genetic correlation between PTSD and agreeableness and a stronger, positive genetic correlation with suicide attempts relative to AUT and ADHD.*

*The Internalizing disorders factor exhibited negative genetic associations with age at menopause, EA, and positive associations with various adverse health outcomes (e.g., asthma, back pain, coronary artery disease). Phenotypes with disorder-specific associations included socioeconomic phenotypes (e.g., owning a house outright), which tended to exhibit slightly stronger negative genetic associations with MDD than with ANX. In addition, we observed a disorder-specific association with neuroticism, where ANX was estimated to have a stronger, positive genetic correlation relative to MDD."*

*We also now write in the Discussion section (page 22):*

*"In line with SNP-level findings noted above, numerous biobehavioral traits also differed in their genetic correlations with AUT to the point where its disorder-specific etiology must diverge substantially from those of the other disorders loading on this factor. Consistent with phenotypic findings[49] and conceptualizations[50] that posit cognitive deficits as a central distinguishing factor across SCZ and BIP, we observe distinct genetic associations with cognitive outcomes, with BIP associated with better*

*outcomes relative to SCZ. Within the personality domain, neuroticism—a construct commonly observed to be both phenotypically[51] and genetically[52] associated across internalizing disorders—showed a stronger association with ANX over MDD. As many of these external traits and the disorders are multi-faceted in nature, it will be important for future work to obtain finer-grained phenotypes to better define the boundaries of these findings. Indeed, recent work using Genomic SEM found that ANX and MDD may share unique genetic associations with specific facets of neuroticism.[53]"*

**6. Accelerometer data: I am not sure what we've learnt from this. The authors also don't discuss this at all later. If the authors have insights, please add to discussions, otherwise I am unsure why this analysis is in the paper.**

*Dysregulation in daily patterns of movements have been posited to reflect a general risk pathway for psychiatric disease. As the accelerometer results show both divergent patterns of findings across the factors, and convergent patterns for the disorders within a factor, this provides substantial evidence for the validity of the factor model. We clarify this point in a revised Discussion section as below (page 22):*

*"At the biobehavioral level, the pattern of associations with external correlates was informative with respect to the shared and distinct characteristics across the disorders. For example, the accelerometer results displayed both divergent patterns of findings across the factors, and convergent patterns for the disorders within a factor. This provides evidence for both the validity and utility of the genetic factor model for characterizing genetic associations with basic aspects of everyday functioning that may be, at face, relatively distal from the biological mechanisms of the disorders themselves."*

**7. Stratified Genomic SEM: the authors wrote that "Stratified Genomic SEM models that allowed variances of the common genetic factors, and disorder-specific effects, to vary across annotations to examine whether the degree of risk sharing and differentiation is enriched across disorders". Does this mean the method is able to identify the degrees of enrichment in variance attributable to common genetic factors, and disorder specific effects, respectively? From Figure 5 it seems the authors have obtained the enrichment for annotations for the factors identified through previous analyses rather than the individual diseases - taking this as the enrichment results for common genetic factors, where are the results for the disease specific effects? Are those estimated? If not, I can't tell what the difference of this analysis is from S-LDSC on the factors, and I would find the description misleading. Please clarify. Further, in terms of the results, as the effective sample size for each factor used in the enrichment analysis are different, the power to identify significant enrichments for each factor would be different. Given the authors are using summary statistics for all analyses and can't down-sample, please discuss how difference in power affect results of this analysis.**

*We are grateful to the reviewer for raising this point. We have now examined enrichment of the residual variance in both the psychiatric factors in the hierarchical model and for the individual disorders in the correlated factors model. Indeed, Stratified Genomic SEM allows us to separate enrichment of genetic convergence from differentiation across groups of disorders, a question that could not be directly answered using existing univariate and bivariate genetic stratification methods (including inputting summary statistics for the factors into univariate S-LDSC). We report these findings in the Results section (page 14):*

*"We went on to examine enrichment of residual (i.e., unique) variance for the individual disorders in the correlated factors model and the residuals of the psychiatric factors in the hierarchical model (Supplementary Table 10). Results for the individual disorders revealed 17 significant residual enrichment estimates at a Bonferroni corrected threshold. This included 13 significant estimates within conserved annotations (e.g., conserved primate; genomic evolutionary rate profiling [GERP]), consisting of 3 conserved annotations enriched for residual variance in MDD, 1 for AUT, 3 for AN, 1 for SCZ, and 5 for ALCH. We also observed significant enrichment unique to MDD for coding regions and the PI × excitatory dentate gyrus neurons annotation. Finally, results revealed four significant annotations reflecting genes expressed early in brain development, including the fetal female brain DNase annotation for ADHD, the germinal matrix H3K6me3 annotation for BIP, and the fetal male brain H3K4me1 annotation for both SCZ and ALCH.*

*Enrichment of the residuals of the psychiatric factors in the hierarchical model indicated slightly attenuated signal across annotations relative to enrichment from the correlated factors model. This was with the notable exception that the enrichment signal was even stronger in the PI × neuronal annotations when examining the residual variance in the Psychotic disorders factor unique of the three remaining factors. This provides compelling evidence that variants within PI genes expressed in specific hippocampal and prefrontal cortex neuronal cells are distinctly important for genetic overlap between BIP and SCZ. Collectively, these findings highlight the unique ability of Stratified Genomic SEM to unpack the genetic underpinnings of both convergence and divergence across traits of interest in a multivariate space."*

*We will also now note as an interpretive limitation the power differences across the factors in the Discussion (page 24):*

*"The current findings at all levels of analysis (biobehavioral, functional, SNP) should also be interpreted with respect to the power of the individual disorders used to define the factors. In particular, the paucity of GWAS hits and significant enrichment findings for the Compulsive disorders factor should be considered in the context of the relatively low power for the disorders that define this factor. Future analyses may also benefit from evaluating these findings using a set of traits that is balanced with respect to statistical power."*

**8. Multivariate GWAS: QQplots for all GWAS show significant inflation, and though the authors have found that the intercept from LDSC were close to 1. However, an intercept close to 1 only suggest that the inflation wasn't likely caused by population stratification, and cannot account for inflation due to all statistical errors. Mostly when inflation was due only to polygenicity and not other errors, the qqplots would be inflated at the low P values (high log P), but the qqplots shown in Supplementary figure 23 all deviate from the null even at the high P values (low log P), making me very suspicious the tests are not well calibrated. The authors plot two qqplots in Supplementary figure 23, corresponding to figure 6 in the main text - one for the factor GWAS (blue), and the other for Qtrait tests (magenta). For the Qtrait P values, how are they calculated (I already asked above), are they well calibrated, and how did the authors check for calibration of these p values? I looked into the original genomicSEM paper in addition to this paper, and didn't see any tests for calibration of Qtrait. Similarly, can the authors show that the factors capture the shared effects of**

**triats they load onto, and do not inflate their effects? It would be hard to trust these values without demonstration they are calibrated.**

*We appreciate the concern raised by the reviewer and to this end have added a Supplemental Figure (Figure S32b) that depicts the QQ-plots for the factors along with the univariate GWAS for the traits that load onto the factors. As can be seen in this figure, these results show even larger inflation for some of the disorders relative to the factor. This strongly suggests that the concern raised is not unique to Genomic SEM, but rather is a reflection of the high statistical power of the included univariate GWAS statistics on which Genomic SEM is based. The data have been carefully QCd for a variety of confounds (including technical confounds such as batch effects) for the original univariate studies from which we draw and integrate summary data, with the general consensus in the field being that the early lift-off of QQ plots is largely attributable to the polygenic nature of psychiatric traits. We have also updated the Method section to note some specific findings from the original Genomic SEM publication (Grotzinger et al., 2019) that speak to the reviewer's concern about Genomic SEM capturing shared signal (page 25):*

*"Validation of Genomic SEM in Grotzinger et al. (2019)[14] demonstrated that the framework produces unbiased standard errors, appropriately accounts for sample overlap in multivariate GWAS, and produces accurate point estimates for different population generating models. In addition, polygenic scores derived from Genomic SEM summary statistics were found to better predict the individual traits that define the factor than PGS constructed from the summary statistics for the individual traits. As part of the current analyses, we sought to further validate Genomic SEM via a series of simulations based directly on the factor structure identified here and additionally benchmark Genomic SEM against existing multivariate methods."*

*The reviewer also requests that we provide demonstration of the calibration of the Q statistic. This was originally presented as a test of heterogeneity of SNP effects, and we have revised the current manuscript to make clear that the mathematical properties of $Q_{Trait}$ are identical to those of $Q_{SNP}$, as noted in response to the reviewer's third point. We also now include simulations to report how Q performs under a variety of population generating models and demonstrate that Q is well-calibrated, with power increasing as a function of the population generating model increasingly deviating from the identified factor structure (see response to Reviewer 2 Comment 10 below for full details of simulation results).*

**Minor**

**1. Figure S8 has nano-sized fonts, really difficult to read, please consider breaking up into multiple pages like Figures S4, S7, S9 etc**

*We have increased the font size and broken up this figure across multiple panels as requested.*

**2. I am not opposed to polar plots but think they are unnecessarily hard to read. Normal bar plots would do the jobs just as nicely.**

*As requested, we have updated all polar plots to be bar plots in the Online Supplement.*

**3. For most of the supplemental figures, perhaps it's much easier to represent these as tables.**

*All supplemental figures now have corresponding supplementary tables. In particular, we highlight that the information depicted in Supplementary Figures 6 and 7 of the bar plots of the genetic correlations for the individual traits is now additionally reported in Supplementary Tables 4 and 5. In addition, the results from Supplementary Figure 11 depicting the Stratified Genomic SEM factor model simulation results are now additionally reported in Supplementary Table 8.*

**Reviewer #2:**
**Remarks to the Author:**

**The authors use genomic structural equation modeling to infer 4 factors underlying 11 psychiatric disorders and draw inferences about the genetic architecture of these 4 factors. They then conduct a GWAS of the 4 factors, identifying 152 loci including 20 novel loci.**

**Section identifying the 4 factors:**

**(1) This is only a mild extension of the results of Lee et al. 2019 Cell (ref. 13), who analyzed 8 of these psychiatric disorders and identified 3 of these factors. The authors are transparent about this. Thus, this section should be made much shorter. Confirmatory factor analyses and other subtle details can be moved entirely to the supplement. The comments about Cai et al. 2020 Nat Genet (ref. 28), though appropriate, can be moved to the Discussion.**

*We note that by expanding the range of disorders, we have elaborated the original 3 factor model reported in Lee et al. (2019) to now include an important fourth factor reflecting internalizing disorders. This factor, which includes major depressive disorder (which had previously clustered with the psychotic disorders), represents an important dimension of variation of major clinical significance. We also highlight that the original PGC Cross-disorder paper (CDG1, 2013) consisted of five major disorders. The major update by Lee et al. (2019) for the second major PGC Cross-disorder effort (CDG2) then included three additional disorders, for a total of eight disorders. Similarly, we add an additional three disorders for a total of 11 disorders, and substantially update many of the sample sizes relative to those from the Lee et al. (2019) publication. This said, we have shortened this section and moved some details of the model fitting procedure to the Method section and the comments about the Cai et al. 2020 paper to the Discussion as requested per below (page 24):*

*"Moreover, our results may have been influenced by the phenotyping and case-ascertainment methods used. Cai et al. (2020)[66] have specifically reported that psychiatric phenotypes derived using minimal phenotyping (defined as "individuals' self-reported symptoms, help seeking, diagnoses or medication") may produce GWAS signals of low specificity. Although our sensitivity analyses suggested minimal differences when excluding GWAS that used self-report cohorts this issue should continue to be explored in future work."*

**(2) The p-factor model seems to be of low interest. It has greater complexity, and has close to zero genetic associations in the GWAS part of the paper. Either all content on the p-factor model should be moved to the supplement, or considerable additional justification is needed as to why it is important in this paper.**

*We agree with the reviewer that our results with respect to the p-factor (e.g., close to zero GWAS genetics associations, no meaningful mediation of the relationship between external traits and psychopathology as measured by $Q_{Trait}$) indicate that the p-factor is of little theoretical or pragmatic utility. If these results were motivated by a novel hypothesis on our part, we would agree that the fact that they were unsupported would limit their interest. However, the p-factor is a construct that is widely cited within the literature to the extent that it is often implicitly treated as "real" or at least meaningful. For example, the 2014 Caspi et al. paper (DOI: 10.1177/2167702613497473) naming the p-factor has been cited over 1,400 times according to Google scholar. Thus, we believe that our lack of support for the p-factor, in the context of the very high general interest in this transdiagnostic construct, is strong rationale for including the results in the main text.*

**(3) Minor comments: The authors should include a main Table that includes all of the information in Table 1 of Lee et al. 2019 Cell (ref. 13). Also, Figure 1C and Figure 1D should be moved to the supplement. (Not clear what Figure 1D refers to, as the text does not mention a "bifactor model".) The text fragment "prior literature indicating a high-order transdiagnostic 'p-factor'" requires a reference, perhaps ref. 6 or ref. 6-8 as cited previously.**

*We have added the sample sizes, population prevalences, and estimated heritabilities to Table 1 to mirror the corresponding Table from Lee et al. (2019). We have also included the requested references (i.e., Caspi et al., 2014; Lahey et al., 2012; Petterson et al., 2016) for the noted text fragment. In addition, we now signal to the reader in the factor analysis section of the results that the bifactor model in Figure 1D is discussed later on. In line with our response to the reviewer's second point directly above, we have made the case to retain Figure 1C and 1D.*

**Section on genetic correlation of the 4 factors with external traits:**

**(4) Qtrait is not a compelling metric, both because statistical significance is sample size dependent, and because the null hypothesis that the genetic correlation between an external trait and the 11 psychiatric disorders operates entirely through one of the factors is not a plausible null hypothesis, such that its violation is not meaningful. It might be more interesting to assess what proportion of the genetic correlation between an external trait and the 11 psychiatric disorders operates through one of the factors (or all 4 factors). Also, Figure 2 should be moved to the supplement.**

*We agree that the pattern of associations with external traits is unlikely to perfectly map onto the estimated factor loadings, in the same way that most effects in the population are unlikely to be exactly 0. As we identify only a subset of traits and SNPs as significant for Q, we believe that this test is not calibrated to identify an abundance of significant findings. Indeed, we now show in the new simulations added for this revision that when the pattern of associations with an external correlate deviate only slightly from the model, Q is not prone to being significant, but is well-powered when SNP associations strongly deviate from the model expectations. We have made the null of the Q metrics more explicit in the updated Method section in addition to re-iterating the explication from the supplement of de la Fuente et al. (2020), that makes clear that because Q is a hypothesis test regarding the extent to which regression coefficients (and not Z statistics) conform to a factor loading pattern. We additionally note that Q is not biased by differential power across traits (see response to Reviewer 1 Point 3 for the specific updates to*

*the Method section). As both reviewers raised questions about the Q metric and how it was calculated, we have retained Figure 2 in order to aide readers in the interpretation and calculation of this test statistic, and we now include simulations showing that Q appropriately increases in power as the population generating model reflects a corresponding increase in deviation from the factor structure specified in Genomic SEM (see response to Reviewer 2 point 10 below for simulation details and results).*

**(5) Some of the genetic correlations of the 4 factors with external traits are interesting (particularly for the accelerometer traits), but it is unclear whether the insights gained are more impactful than the insights gained by assessing (only) the genetic correlations of the 11 psychiatric diseases with external traits (for example, in Figure 4, results for 4 factors look similar to results for corresponding psychiatric diseases). It would be important to discuss this.**

*The reviewer's observation that many of the genetic correlations with the 11 psychiatric diseases appear similar to the correlations identified for the 4 factors speaks to the utility of using Genomic SEM to understand these patterns of correlations. We clarify the value of testing the validity of the factors using the evidence from external patterns of genetic correlation. This is particularly compelling because the factor model itself was developed on the basis of internal patterns of genetic correlations among the disorders. That is, given similar patterns observed in the genetic correlation matrices, Genomic SEM allows us to formally quantify the genetic correlation between psychiatric traits and external indicators and to test using* $Q_{Trait}$ *whether these patterns are significantly different across the psychiatric traits. We specifically write in the Results section (page 7):*

*"A factor model implies a specific causal model in which the factors influence their indicators, i.e. the Compulsive Disorders, Psychotic Disorders, Neurodevelopmental Disorders, and Internalizing Disorders factors each influence the subset of individual psychiatric disorders that load on them. Therefore, the identified factor structures also imply a patterning of certain genetic relationships between external traits and the individual disorders. The degree to which the observed genetic correlation between traits and the psychiatric disorders respect the relationships implied by the factors can be viewed as a validation, or rejection, of the factor structure at one level of analysis."*

*We additionally highlight in the Discussion (page 22):*

*"At the biobehavioral level, the pattern of associations with external correlates was informative with respect to the shared and distinct characteristics across the disorders. For example, the accelerometer results displayed both divergent patterns of findings across the factors, and convergent patterns for the disorders within a factor. This provides evidence for both the validity and utility of the genetic factor model for characterizing genetic associations with basic aspects of everyday functioning that may be, at face, relatively distal from the biological mechanisms of the disorders themselves."*

**Section on genetic enrichment of the 4 factors in annotations:**

**(6) "Stratified Genomic SEM allows us to ask whether pleiotropic loci are enriched within particular annotations": I agree with this comment. However, why not just stratify genetic covariance by annotation, which also assesses whether pleiotropic loci are enriched within particular annotations? (The authors actually develop this approach by developing stratified**

**multivariate S-LDSC as the first step of their Stratified Genomic SEM method, although this is not made clear until the Methods section. Also, Lu et al. 2017 Am J Hum Genet is a published method for stratifying genetic covariance by annotation that should be cited and discussed.) It is unclear whether the insights gained from the factor approach are more impactful than the insights gained by (only) stratifying genetic covariance by annotation. It would be important to discuss this. Specifically, are there any annotations for which the genetic enrichment for one of the 4 factors is significantly larger in magnitude than the genetic enrichment for any of the 11 psychiatric disorders?**

*There are a number of advantages to examining enrichment at the level of the factors, rather than at the level of individual variables or pairwise combinations of variables. Enrichment of genetic covariance provides information about the gene categories in which directionally concordant, pleiotropic signal, as opposed to heritability itself, is disproportionately localized. However, when data become multivariate, pairs of variables rapidly expand, which renders examination of enrichment itself high dimensional. Genomic SEM resolves this issue by allowing for enrichment to be examined at general dimensions of genetic variation spanning large constellations of (rather than simply pairs of) phenotypes. In response to Reviewer 1 Comment 7, we now further capitalize on the multivariate nature of Stratified Genomic SSEM to examine enrichment of the residual genetic variance of the indicators. This allows us to ask whether there is enrichment of signal that is unique to one of the indicators, which would not be possible by performing standard S-LDSC on the factors or examining stratified genetic covariance in isolation. We have updated the Results section to be more explicit about the validation of multivariate S-LDSC as part of the development of Stratified Genomic SEM (page 12):*

*"Overview and Validation via Simulation. We developed Stratified Genomic SEM to allow the basic principles of Genomic SEM to be applied to genetic covariance matrices estimated within different gene sets and categories (Method). These gene sets and categories, collectively referred to as annotations, can be constructed based on a variety of sources, such as collateral gene expression data obtained from single-cell RNA sequencing. Such an analysis goes beyond methods such as Stratified LDSC (S-LDSC)[33] that estimate enrichment of heritability for particular traits within functional annotations. Rather, Stratified Genomic SEM utilizes a multivariate framework to ask whether shared and unique genetic signal across a set of traits is enriched within particular annotations. Enrichment is defined as the ratio of the proportion of genome-wide risk sharing indexed by the annotation to that annotation's size as a proportion of the genome (Method). The null, corresponding to no enrichment, is a ratio of 1.0, with values above 1.0 indicating enriched signal within a functional annotation.*

*In order to validate the key statistical properties of Stratified Genomic SEM, we began by simulating genetically correlated phenotypes that were enriched in six annotations. We then show that our multivariate extension of S-LDSC produces accurate estimates of stratified genetic covariance along with unbiased standard errors (Supplementary Figures 8-10; Supplementary Tables 7-9). Finally, we demonstrate that these stratified genetic covariance matrices can be used as input to Stratified Genomic SEM to produce unbiased factor loadings, and unbiased standard errors (Supplementary Figure 11)."*

*We have also added the noted reference in the Discussion section as requested and, in line with our response to the reviewer, highlight the benefits of Stratified Genomic SEM (page 23):*

*"In order to identify gene sets and categories in which shared and unique genetic signal for multiple disorders is disproportionally localized, we developed and validated both a multivariate extension of S-LDSC and Stratified Genomic SEM. We note that genetic covariance analyzer (GNOVA) is an existing bivariate framework that utilizes method of moments to also estimate stratified genetic covariance using GWAS summary statistics.[56] This is expected to produce similar estimates to our multivariate extension of S-LDSC in the bivariate context. However, Stratified Genomic SEM is the first multivariate framework capable of examining enrichment across larger constellations of correlated traits. In the first application of this method, we use Stratified Genomic SEM to examine enrichment for 168 annotations across four factors of psychiatrics risk, including 29 annotations representing protein-truncating variant (PTV)-intolerant (PI) genes, genes expressed in the human brain cells in the hippocampus and prefrontal cortex, and their intersection."*

**(7) Minor comment: the term "genetic correlation" usually refers to a relationship that has consistent sign across the genome. The term "pleiotropic" usually refers to a relationship that may have varying sign across the genome. This paper seems to focus on relationships that have consistent sign across the genome. Thus, either the term "genetic correlation" (or covariance) should be consistently used, or at least the terminology (including the use of the term "pleiotropic") should be carefully defined/explained.**

*We have updated the text to be clearer when we are referring specifically to the case of unidirectional pleiotropy or pleiotropy more generally. For example, we write in the section on unstructured multivariate GWAS (page 16):*

*"Moreover, 7 hits were entirely novel in that they were not in LD with any previously discovered hits in the GWAS catalogue. For comparative purposes, we consider overlap with the 109 pleiotropic (i.e., associated with more than one disorder irrespective of directionality) and 146 total hits from PGC-CDG2[13] given both overlapping datasets and research questions."*

*We additionally note in the subsequent section on structured multivariate GWAS findings (page 16):*

*"Taken together, loci that are significant for the factor, and not for that factor's $Q_{SNP}$ metric, are generally expected to be unidirectionally pleiotropic at the level of the disorders that define that factor."*

**Unstructured multivariate GWAS and structured multivariate GWAS sections:**

**(8) The "unstructured" GWAS (which does not make use of the 4 factors) identified 184 loci including 39 novel loci. The "structured" GWAS (which does make use of the 4 factors) identified 152 loci including 20 novel loci. Why is the structured GWAS preferred? Are its results more actionable, and if so why?**

*We thank the reviewer for pointing out this interpretive consideration. We have updated the Results section to note on page 15:*

*"For the purposes of the current investigation, the factor model and unstructured model are each informative in their own right. The unstructured model is particularly well-suited when the aim is to*

*identify an exhaustive set of SNPs relevant to psychiatric risk, but does little to elucidate the specific patterning of associations. In contrast, the factor model allows us to systematically probe the genetic underpinnings of convergence and divergence across clusters of psychiatric disorders."*

**(9) How do the unstructured GWAS method and the structured GWAS method compare to the considerable literature on multi-trait GWAS methods that has been published, including Nieuwboer et al. 2016 Am J Hum Genet (not cited, same last author as this paper) and Turley et al. 2018 Nat Genet (not cited) and Lee et al. 2019 Cell (ref. 13) and Grotzinger et al. 2019 Nat Hum Behav (ref. 44; same first author as this paper)? In particular, is the structured GWAS method proposed in this paper identical to the structured GWAS method proposed by ref. 44? Although ref. 44 is cited in the structured GWAS method, this is not explicitly clear.**

*The reviewer is correct that the structured GWAS is estimated using Genomic SEM, which we introduced in our 2019 paper in Nat Hum Behavior. We have updated the terminology to refer to these as unstructured and factor (previously referred to as structured) models as we believe this more clearly maps onto the form of multivariate GWAS estimated. We also now highlight in the updated Results section on factor model multivariate GWAS (page 15):*

*"For context, we note that a multivariate GWAS estimated using a factor model is also what was presented in the original Genomic SEM publication.[14]"*

*We refrain from citing the work by Nieuwboer et al. 2016 as this is not actively being developed to be used for structure multivariate analyses, and we have previously shown in Grotzinger et al. (2019) that many of its relevant capabilities are subsumed by the Genomic SEM framework. Please see our response to the next comment for further detail on results comparing Genomic SEM to existing multi-trait methods.*

**(10) For these sections to have high methodological impact, it would be critically important to add simulations comparing the calibration and power of the proposed structured GWAS method to other multi-trait methods in simulated data, and explain (with justification) in which scenarios the method provides an increase in power.**

*As one of the major updates to the resubmission, we include simulations to benchmark the performance of Genomic SEM under nine different population generating conditions. We specifically compare Genomic SEM to the algorithms for three main, multi-trait methods: MTAG, N-GWAMA, and MA-GWAMA. We note that the present submission focuses specifically on examining the multivariate genetic architecture of psychiatric disorders, including identifying SNPs that both operate through, and diverge from, this structure. Simulation results reported below indicate that Genomic SEM is particularly well-calibrated for investigating these research questions and that existing alternatives are neither calibrated nor designed for addressing these questions. We specifically write in the updated Results section (page 14):*

***"Multivariate GWAS***

***Simulations.*** *We conducted a series of simulations to further validate the calibration of Genomic SEM for multivariate GWAS in the specific context of the analyses presented here. The population-generating model for the first scenario was that implied by the correlated factors model, in which a SNP is directly*

*associated with all four factors, but not uniquely associated with any of the 11 individual disorders. This scenario represents the best-case scenario for the factor model, in that all SNP effects are specified to occur on the factors. Subsequent scenarios specified population-generating models in which SNP effects increasingly deviated from that implied by the factor structure. For each simulation, we used Genomic SEM to estimate factor-specific SNP effects and factor-specific indices of heterogeneity, as indexed by $Q_{SNP}$.[14] $Q_{SNP}$ indexes violation of the null hypothesis that the SNP acts on the individual disorders entirely via the factor on which they load (Figure 3; see* ***Method****). As expected, simulation results revealed that the power to detect multivariate SNP effects and $Q_{SNP}$ decreased and increased, respectively, as population SNP effects increasingly deviated from those implied by the factor structure (Supplementary Figures 25-28, Supplementary Table 11). These simulations additionally illustrated that that SNP effects on factors, as estimated with Genomic SEM, are not simply the reflection of the most high-powered univariate GWAS that defines the factor, that there is null signal when the population of SNP effects is set to 0, and that power for $Q_{SNP}$ is particularly high when there are directionally discordant SNP effects across the factor indicators.*

*We went on to benchmark these simulation results against those obtained from alternative multivariate methods. We specifically chose three methods—Multi-trait Analysis of GWAS (MTAG),[44] Model Averaging Genome-wide Association Meta-analysis (MA-GWAMA), and N-weighted Multivariate GWAMA (N-GWAMA)[45]—that also use both summary statistics and take into account genetic covariance across phenotypes. In addition, we examined the performance of multivariate GWAS in Genomic SEM when specified as an unstructured model that computes an omnibus index of association across all 11 disorders. Unstructured model results were obtained by comparing a maximally complex model in which the SNP is allowed to have direct regression relations with each of the 11 disorders against a null model in which the SNP is associated with none of the disorders. This omnibus test is $\chi^2$ distributed with 11 df, and quantifies evidence for an overall effect of the SNP on any subset of the disorders, irrespective of the patterning or directionality of the effects. We refer to this as an unstructured model because the tested model freely estimates as many SNP regressions as there are disorders. This is in contrast to the multivariate GWAS specified as a factor model discussed initially that estimates SNP effects on the factors, as this defines a specific pattern, or structure, of the relationship between the SNP and the 11 disorders. For context, we note that a multivariate GWAS estimated using a factor model is also what was presented in the original Genomic SEM publication.[14]*

*Compared to the other methods, simulation results revealed that the factor model is relatively better powered when the patterning of SNP effects on the disorders is consistent with that implied by the model and least powered as the population deviates substantially from that implied by the factor model (Supplementary Figures 29-31; Supplementary Table 11). In addition, we find that all methods evince null signal in the presence of null SNP effects in the population-generating model. An unstructured model had the expected property of being better powered than alternative methods when the SNP was associated with only a subset of traits or had directionally opposing effects across traits. Rather than speaking to the universal superiority of any individual method, simulation results indicate that the preferred method will depend on the specific research questions. For the purposes of the current investigation, the factor model and unstructured model are each informative in their own right. The unstructured model is particularly well-suited when the aim is to identify an exhaustive set of SNPs relevant to psychiatric risk, but does little to elucidate the specific patterning of associations. In contrast, the factor model allows us to systematically probe the genetic underpinnings of convergence and divergence across clusters of psychiatric disorders."*

*We provide additional detail in the Method section, writing (page 25):*

*"**Multivariate GWAS Simulations***

*Simulation Procedure. In order to examine the calibration of Genomic SEM for multivariate GWAS, we began by estimating the model implied genetic covariance matrix for a model in which rs9314056—a hit for the Internalizing disorders factor and a univariate hit for MDD and ANX—was specified to predict the four factors from the correlated factors model. Nine different versions of this genetic covariance matrix were used to form population generating covariance matrices from which individual covariance matrices were simulated using the rmvrnorm function in the rockchalk R package. The observed sampling covariance matrix (V) was used for sampling from the population matrices, and was subsequently paired with each simulated genetic covariance matrix when estimating the model in Genomic SEM. As the V matrix includes squared SEs on the diagonal, simulated parameters (e.g., the genetic covariance between MDD and ANX; the association between the SNP and PTSD, etc.) were therefore specified to have the same precision as in the observed data. This has the intended consequence that the simulations reflect the empirical data scenario wherein certain associations are estimated with greater precision, as will often be the case when the contributing univariate GWAS was estimated using a larger participant sample. We have therefore endeavored to conduct a series of simulations that are both directly relevant to the current analyses and more broadly reflect the realistic scenario of differentially powered GWAS entered into the same multivariate framework.*

*Genetic covariance matrices were sampled 250 times for nine different population generating scenarios, for a total of 2,250 simulations. These nine scenarios consisted of: Scenario 1 in which the model implied matrix was unchanged; Scenario 2 in which the covariance between the SNP and ALCH was set to 0; Scenario 3 in which the covariance between the SNP and PTSD was set to 0; Scenario 4 in which the covariance between the SNP and ANX was set to 0; Scenario 5 in which the covariance between the SNP and MDD was set to 0; Scenario 6 in which the covariance between the SNP and PTSD, ALCH, and ANX was set to 0; Scenario 7 in which the covariance between the SNP and MDD, PTSD, ALCH, and ANX was set to 0; Scenario 8 in which the covariance between the SNP and all 11 psychiatric traits was set to 0; and Scenario 9 in which the direction of the covariance between the SNP and ANX and ALCH was reversed (i.e., multiplied by -1). These nine scenarios were chosen to reflect varying degrees of conformity to the Internalizing disorders factor structure, with Scenario 1 exactly matching the model and Scenario 9 reflecting the most extreme deviation from the model wherein the SNP has directionally opposing effects on ALCH and ANX. We include Scenario 8 in addition to Scenario 7 as the estimated SNP effects for the Internalizing disorders factor may include some minimal genetic signal from the broader correlated factors model. Note that none of the subsequent models estimated in Genomic SEM fixed the relationship between a psychiatric trait/factor and SNP to 0, nor were the simulated covariance matrices likely to produce a SNP-trait relationship at exactly 0. Rather, SNP-trait associations were only set at 0 in the generating population.*

*In the sections below, we first compare results across the nine different population generating scenarios for a factor model multivariate GWAS in Genomic SEM in which the SNP effect was specified to predict the four factors from the correlated factors model. We subsequently compare these results to those from an unstructured GWAS (discussed further below) in Genomic SEM that seeks to provide an exhaustive list of SNPs relevant to the traits of interest. This is in contrast to the factor model results that estimates SNP effects specified to operate via the structure of the factors. We additionally consider results*

*across three, separate multivariate GWAS methods: MTAG,[69] N-GWAMA,[45] and MA-GWAMA,[45] also discussed further below.*

*Factor Model. We first examined the distribution of estimated SNP effects and $Q_{SNP}$ specific estimates for the Internalizing disorders factor in Genomic SEM across the nine scenarios. As expected, the distribution of estimated SNP effects revealed the strongest signal for Scenario 1 in which the population exactly matched the factor model (Supplementary Figure 25), with all 250 runs producing genome-wide significant hits for the Internalizing disorders factor (i.e., no false negatives) and an average p-value for the estimated SNP effect on the factor of 2.51E-10 (Supplementary Table 11). The signal was also comparable for Scenarios 2 and 3 where the SNP association with the two disorders with the smallest factor loadings, ALCH and PTSD, was 0 in the population (Supplementary Figures 1 and 3). This was followed by reduced signal when the SNP with ANX association was 0 (Scenario 4), the SNP association with PTSD, ANX and ALCH was set to 0 (Scenario 6), and the SNP association with MDD was set to 0 (Scenario 5).*

*A particular concern for the Internalizing disorders factor may be that the larger sample size for MDD relative to the other three disorders that load on this factor results in estimated factor SNP effects that merely recapitulate the signal for MDD. Scenario 6 was designed to test this concern. As can be seen in the distribution of effects (Supplementary Figures 25 and 27) there is a marked downshift in the signal for this scenario when all SNP associations with Internalizing indicators except MDD were set to 0 in the population. This demonstrates that while SNP associations with a factor will certainly be more influenced by a better powered factor indicator that also has a larger factor loading, that the signal is not strictly dominated by this indicator. As would also be expected, the signal was particularly attenuated for Scenario 9 when the direction of the SNP association with ALCH and ANX was reversed, and was the weakest for Scenarios 7 and 8 in which the SNP association with the four Internalizing factor disorders and all 11 disorders were set to 0 in the population. Moreover, there were no factor hits (i.e., no false positives) in the latter two scenarios, and all SNPs in Scenario 9 were estimated as hits for $Q_{SNP}$.*

*The trends for $Q_{SNP}$ were also in the expected directions. More specifically, there was a clear null signal for $Q_{SNP}$ for Scenario 1 for which the model matched the population (Supplementary Figure 28), no $Q_{SNP}$ hits (i.e., no false positives) and an average $Q_{SNP}$ p-value of .560. There was a similar absence of signal for Scenarios 2 and 3, also with no $Q_{SNP}$ hits and no deviation from the expected p-values in the QQ-plot. In addition, there was very little signal for Scenarios 7 and 8 where trait and SNP associations were at 0. This is also expected, as estimated SNP associations that are consistently near 0 across indicators that load on the same factor are, in fact, not hugely discrepant from the factor model. $Q_{SNP}$ signal increased for the scenarios that more strongly deviated from the structure, in which ANX, MDD, or PTSD, ALCH and ANX were 0 in the generating population. The signal was by far the largest for Scenario 9 (Supplementary Figure 28) in which the directionality of the SNP effect was reversed for ANX and ALCH, with 100% of the 250 runs estimated as genome-wide significant $Q_{SNP}$ hits. This is the scenario that deviated strongest from the factor model and, in line with observation, is expected to pick up on the largest $Q_{SNP}$ signal.*

*Comparing Genomic SEM to Other Multivariate Methods. In the absence of other summary statistics based SEM methods, we sought to perform a comparison of Genomic SEM to three of the most closely related multivariate methods: MTAG,[69] N-GWAMA,[45] and MA-GWAMA.[45] These methods were considered most similar to Genomic SEM in that they also account for unknown degrees of sample overlap via the bivariate LDSC intercept and produce results by statistically incorporating the estimated*

*genetic covariance across included traits. We additionally compare results to an unstructured model in Genomic SEM that seeks to identify an exhaustive set of SNPs relevant to the traits of interest, irrespective of directionality. MTAG, N-GWAMA, and MA-GWAMA, utilized only the four internalizing disorder indicators (ANX, PTSD, ALCH, MDD) to mirror the factor model simulation results presented above for the Internalizing disorders factor. Before comparing simulation results across methods, we first provide a brief overview of each method and how results were produced using our simulation procedure. We refer to the reader to the original articles for further details on estimation procedures and statistical properties for each method.*

***Unstructured Model.*** *We estimate SNP effects via an unstructured model in Genomic SEM by calculating a model $\chi^2$ difference test for a model in which the SNP is allowed to have direct regression relations with each of the 11 disorders (i.e., a fully saturated model) against a null model in which the SNP is associated with none of the disorders. This omnibus test is $\chi^2$ distributed with 11 df, and quantifies evidence for an overall effect of the SNP on any subset of the disorders, irrespective of the patterning or directionality of the effects. These models do not include any higher order factors and are meant to provide an exhaustive list of SNPs associated with included traits. All 11 disorders were included for the unstructured models, despite choosing simulation parameters for a SNP that is specifically relevant to the Internalizing disorder factor and indicators. By including all 11 disorders, the simulations mirror the real data analyses conducted and provide a more conservative test of the unstructured GWAS approach. That is, if the goal is to identify a comprehensive set of associated SNPs, it is most informative to examine the performance of the unstructured models for scenarios in which the SNP affects only a subset of the included traits.*

***Multi-trait Analysis of GWAS (MTAG).***[69] *MTAG works by leveraging the shared genetic information across traits, as indexed by the LDSC genetic covariance, to increase power for a particular trait. The MTAG model was specified in Genomic SEM in order to directly use the simulated genetic covariance matrices for analyses. We have shown previously that MTAG specified in Genomic SEM produces estimates that are correlated at > .99 with summary statistics produced from the original MTAG software.*[14] *We specified MDD to be the MTAG "target" and PTSD, ADHD, and ANX as the secondary traits used to boost signal; a schematic of the MTAG model for MDD as estimated in Genomic SEM is depicted in Supplementary Figure 29.*

***Model Averaging GWAMA (MA-GWAMA).***[45] *MA-GWAMA functions by first estimating a manifold of models that specify the simple regression relationship between the SNP and a set of traits using distinct design matrices, X. Mirroring the original MA-GWAMA software, X is dichotomously coded (0, 1) such that the models allow for the existence of two distinct genetic effects across two traits or subsets of traits. The estimates from these models are then aggregated using weights derived from the fit of the model, as indexed using $AIC_C$. In order to mirror the format of results expected by the software, all simulated SNP-phenotype covariances and corresponding standard errors were transformed into SNP-phenotype regressions using the simulated SNP variance. As with MTAG, we report MA-GWAMA results for MDD from models that additionally included PTSD, ADHD, and ANX.*

***N-Weighted Multivariate GWAMA (N-GWAMA).***[45] *N-GWAMA produces a single multivariate test statistic that is computed as the weighted sum of test-statistics taking into account both sample*

*overlap and the genetic covariance across included traits. Reported simulation results then reflect a weighted aggregate across MDD, PTSD, ADHD, and ANX, as opposed to an updated test statistic for MDD as in the case of MA-GWAMA and MTAG. The SNP-phenotype covariances were also transformed to SNP-phenotype regressions to mirror the expected format of results for the N-GWAMA software.*

*We highlight results for a few key scenarios here. For Scenario 1, in which the generating population matched the specified model, the factor model in Genomic SEM was slightly better powered than the other methods (Supplementary Figures 30-31; Supplementary Table 11). For Scenario 5, in which the population generating SNP and MDD association was 0, the unstructured and factor models were generally better powered than the remaining three methods. Conversely, for Scenario 6 in which the population generating SNP association was 0 for all Internalizing traits except MDD, the signal was the most reduced for the factor model. This pattern of results is consistent with the analytic goals of each individual method. For Scenarios 7 and 8, in which the population generating SNP associations were zero, results revealed similarly null signals across all methods. Finally, in Scenario 9 in which the SNP association with ANX and ALCH was directionally reversed, the factor model and unstructured model showed the weakest and strongest signal, respectively, compared to the other three methods.*

*These results collectively speak to the fact that, relative to other multivariate methods, the multivariate GWAS signal for the factor model is not dominated by any single trait and is particularly sensitive to distinct patterns of SNP associations across traits. In addition, an unstructured model was especially well-suited for identifying a comprehensive set of SNPs associated with the traits. This does not indicate that Genomic SEM should be universally preferred over other multivariate genomic methods, as many approaches seek to increase signal for a particular target trait. Indeed, consistent with this particular analytic goal the signal was more deflated for the MTAG model and MA-GWAMA relative to Genomic SEM when the population SNP effect was 0 for the target trait, MDD. For the current investigation, the analytic goals reflect identifying SNPs generally associated with psychiatric risk and characterizing the genetic underpinnings of convergence and divergence across clusters of psychiatric disorders. The current simulations indicate that unstructured and factor models are particularly well-suited for these purposes in both an absolute sense and relative to existing alternatives."*

**(11) For these sections to have high biological impact, it would be critically important to add considerable biological interpretation of the novel findings. Also, it would be preferred to empirically replicate novel findings in independent data sets, as an additional validation.**

*We have updated the Results and Discussion to consider the prior literature relevant to the pattern of enrichment findings identified in the current analyses.*

*We write in the Results section (page 14):*

*"Genetic sharing across disorders, as estimated by a higher order p-factor, was enriched in conserved annotations and enrichment increased from low to high MAF alleles (Supplementary Figures 19-24). This indicates that previous reports of similar findings for individual disorders[33,41] may reflect enrichment of variants that are broadly relevant for many disorders. The most enriched annotations for the Neurodevelopmental and Internalizing disorders factors were fetal female brain DNase and fetal male brain H3K4me1, respectively, both of which have been previously reported to be enriched for general liability across psychiatric disorders.[42] For*

*specific tissues, brain regions were generally enriched, as was also observed for other complex traits,[43] but were most enriched for the Psychotic disorders factor."*

*We now write in the Discussion section (page 23):*

*"Excitatory hippocampal CA1[57,58] and CA3[59,60] neurons and GABAergic neurons[61,62] have all been associated with risk for both SCZ and BIP. In line with these findings, we observe that the intersection between PI genes and genes expressed in both excitatory and GABAergic neurons explained an outsized proportion of the genetic variance in the Psychotic disorders factor, which primarily indexes genetic covariance between SCZ and BIP. These results indicate that the noted converging lines of evidence for these disorders in prior univariate work likely reflect shared risk pathways and offer critical insight into increasingly specific classes of genes relevant to the high genetic overlap across SCZ and BIP. Enrichment of variance unique to MDD in excitatory dentate gyrus (DG) neurons is consistent with prior findings on the anti-depressive effects of DG stimulation in mouse models[63] and the observation that anti-depressants increase neurogenesis in this region.[64] Additional enrichment for genetic variance unique to disorders was observed for four early brain development annotations, suggesting that at least part of what differentiates disorders may be present very early in brain development. This inference stands in contrast to etiological frameworks that posit the development of transdiagnostic risk factors prior to the later differentiation of disorder-specific risk. Across both the factors and at the level of genetic variance unique to the disorders, we find that conserved regions are generally enriched. As enrichment in conserved annotations has been previously reported for both psychiatric traits and a host of other complex traits (e.g., cognitive function, anthropometric traits[41,43]), the current findings suggest that these annotations capture a range of both shared and unique signal across many domains of functioning and illness."*

*While we would ideally replicate these findings as the reviewer suggests, many of the included disorders represent the only GWAS of sufficient sample size to be included in analyses such as Stratified Genomic SEM or LDSC more generally. We now note this as a general limitation and avenue for future analyses in the Discussion section (page 24):*

*"Additionally, it was not possible to validate our findings in independent datasets owing to the fact that secondary datasets of sufficient sample size do not yet exist for many of the included disorders. The replicability of these findings will of course be critical to examine in future analyses."*

**(12) QSNP is not a compelling metric, because the null hypothesis that the association between a SNP and a psychiatric disorder operates entirely through one of the factors is not a plausible null hypothesis, such that its violation is not meaningful. It might be more interesting to assess what proportion of the association between a SNP and a psychiatric disorder operates through one of the factors (or all 4 factors).**

*We would point out that many null hypotheses are often implausible, such as any statistical test in which the null reflects a population effect of exactly 0. It is certainly a fair point that the null of a SNP operating entirely through the factors is equally implausible. However, by setting a stringent significance threshold, it is elucidating to identify via Q those SNPs that strongly deviate from the factor model, such as variants in the ADH1B gene for alcohol use phenotypes. This point is something we have made more explicit in an updated Method section that addresses interpretation of the Q metrics (see response to Reviewer 1*

*Comment 3). In addition, we now show in the simulations that were added for the resubmission that Q is well-calibrated and well-suited for the research questions addressed by the present manuscript: it is not particularly sensitive to slight deviations from the model expectations and highly sensitive to more substantial deviations from the model expectations. Please also see our response to Reviewer 2 Comment 4 that offers a similar critique of the $Q_{Trait}$ metric.*

**(13) Minor comments: definitely state in the Results section that 4,775,763 SNPs were tested for association (intersection across 11 psychiatric disorders) and a genome-wide significance threshold of 5e-08 was used. It would be good to justify the significance threshold by citing previous references in which this rather lax threshold was used. Also, "genome-wide S-LDSC matrix": does this refer to using a genetic covariance matrix derived from S-LDSC instead of LDSC? This should be clarified.**

*We have updated the text in the Results section to note the number of tested SNPs and provide the original citation that justifies using 5e-8 as the significance threshold for common variant analyses in European populations (Pe'er et al., 2007), and note that even the most recent PGC GWAS efforts use this same 5e-8 threshold (e.g., Ripke et al., 2020). We write on page 15:*

*"All multivariate GWAS analyses tested for SNP associations for the 4,775,763 SNPs present in the summary statistics across all 11 disorders and used the standard genome-wide significance threshold of p < 5e-8.[46]"*

*We also now clarify regarding the genome-wide S-LDSC matrix (page 16):*

*"Corresponding results from using the genome-wide S-LDSC genetic covariance matrix that includes all SNPs (i.e., the S-LDSC equivalent to the genetic covariance matrix estimated using LDSC) can be found in Supplementary Tables 32-44 and Supplementary Figures 34-35."*

**The last two Results sections add little to the paper, and should be moved to the supplement.**

*As noted in response to Reviewer 1 Comment 2, the p-factor has gained such a wide attention in the psychiatric literature, including model fitting using a bifactor model, we believe that including the bifactor p-factor analyses will be of wide interest to many in the field. We also believe the MR analyses are both informative and of interest with respect to the pattern of findings, and to the extent that they provide the first demonstration of the ability of Genomic SEM to perform MR analyses. We have aimed to make this latter point clear in the Discussion section (page 21):*

*"In some circumstances, genetic correlations across disorders may arise from direct, potentially mutual, causation between the factor or disorder-specific liabilities and one another[54] or reflect causation directly between the symptoms of different disorders.[55] Based on significant locus-specific violations of the four factor model at loci relevant to ALCH, including a locus in the ADH1B gene, we incorporated Mendelian randomization into Genomic SEM models in order to estimate the direct causal effect of ALCH on risk for the other disorders. Both single- and multi-variant MR indicated causal effects of ALCH on MDD and BIP. The capability to combine MR and Genomic SEM in order to simultaneously*

*model latent variables and direct effects between disorders vastly increases the scope of possible models that can be evaluated in future work."*

*Should the editor agree with the reviewer, or feel that the manuscript needs to be shortened, we are happy to move these sections to the Online Supplement, but aim to make the case as per above to retain these sections in the main article.*

## References

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, *381*(9875), 1371-1379.

Grotzinger, A. D., Rhemtulla, M., de Vlaming, R., Ritchie, S. J., Mallard, T. T., Hill, W. D., ... & Koellinger, P. D. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature human behaviour*, *3*(5), 513-525.

Howard, D. M., Adams, M. J., Clarke, T. K., Hafferty, J. D., Gibson, J., Shirali, M., ... & Alloza, C. (2019). Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*, *22*(3), 343-352.

Lee, P. H., Anttila, V., Won, H., Feng, Y. C. A., Rosenthal, J., Zhu, Z., ... & Wang, M. M. J. (2019). Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*, *179*(7), 1469-1482.

McArdle, J. J., & Goldsmith, H. H. (1990). Alternative common factor models for multivariate biometric analyses. *Behavior Genetics*, *20*(5), 569-608.

Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. (2007). Estimation of the Multiple Testing Burden for Genomewide Association Studies of Common Variants. *Nature Precedings*, 1-1.

Ripke, S., Walters, J. T., O'Donovan, M. C., & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*.

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Adams, M. J., ... & Deary, I. J. (2019). Genome-wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. *American Journal of Psychiatry*, *176*(2), 107-118.

Turley, P., Walters, R. K., Maghzian, O., Okbay, A., Lee, J. J., Fontana, M. A., ... & Magnusson, P. (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics*, *50*(2), 229-237.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190-2191.

**Decision Letter, second revision:**

Our ref: NG-A55864R1

28th October 2021

Dear Andrew,

Your revised manuscript "Genetic Architecture of 11 Major Psychiatric Disorders at Biobehavioral, Functional Genomic, and Molecular Genetic Levels of Analysis" (NG-A55864R1) has been seen by the original referees. As you will see from their comments below, they find that the paper has improved in revision, and therefore we will be happy in principle to publish it in Nature Genetics as an Article pending final revisions to satisfy Reviewer #2's remaining points and to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and we will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics. Please do not hesitate to contact me if you have any questions.

Sincerely,
Kyle

Kyle Vogan, PhD
Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665

Reviewer #1 (Remarks to the Author):

I am satisfied with the authors' detailed response and edits to both the manuscript and supplementary

materials. I recommend publication.

Reviewer #2 (Remarks to the Author):

The authors have been responsive to the reviewer comments.

Minor comments:

(1) The simulation results in Figure S30 and Figure S31 are important, but difficult to decipher based on the unconventional styles of (Figure S30) 5 overlaid bar plots (density of -log_10 p-values), such that in the first few panels all the reader can see is a jumble of 5 colors piled on top of each other, and (Figure S31) Q-Q plots, which are not commonly used in non-null simulations and again lead to 5 colored sets of data points piled on top of each other. Far preferable would be 5 visually distinct curves, e.g. 5 curves denoting -log_10 p-value as a function of percentile (0% percentile to 100% percentile). A strong case could be made for including such an improved figure as a main Figure (restricting to a subset of 3-4 panels).

(2) The authors state in their response to comment (7) that they have updated the text to distinguish between directional vs. non-directional pleiotropy. However, this clarification is provided in only some cases; in other cases, the term "pleiotropy" remains ambiguous in this respect. All uses of the term "pleiotropy" throughout the manuscript (including the Abstract) should be clarified. (I continue to find it odd that the authors sometime refer to "genetic correlation" and other times to "unidirectional pleiotropy", which is a less clear way of saying the same thing, but I recognize that the choice of terminology is up to the authors.)

(3) It would have been far preferable for changes to the manuscript to be tracked or indicated in colored font. The editors may wish to consider whether to require this in future resubmissions to this journal.

**Final Decision Letter:**

In reply please quote: NG-A55864R2 Grotzinger

21st March 2022

Dear Andrew,

I am delighted to say that your manuscript "Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic, and molecular genetic levels of analysis" has been accepted for publication in an upcoming issue of Nature Genetics.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Genetics style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any

additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A55864R2) and the name of the journal, which they will need when they contact our Press Office.

Before your paper is published online, we will be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. <a href="https://www.springernature.com/gp/open-research/transformative-journals"> Find out more about Transformative Journals</a>

**Authors may need to take specific actions to achieve <a href="https://www.springernature.com/gp/open-research/funding/policy-compliance-faqs"> compliance</a> with funder and institutional open access mandates.** If your research is supported by a funder that requires immediate open access (e.g. according to <a href="https://www.springernature.com/gp/open-research/plan-s-compliance">Plan S principles</a>), then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms

that the author or any third party may assert apply to any version of the manuscript.

Please note that Nature Research offers an immediate open access option only for papers that were first submitted after 1 January 2021.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at <a href="https://www.nature.com/reprints/author-reprints.html">https://www.nature.com/reprints/author-reprints.html</a>. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to https://protocolexchange.researchsquare.com/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A55864R2). Further information can be found at https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#protocols

Sincerely,
Kyle

Kyle Vogan, PhD
Senior Editor
Nature Genetics
https://orcid.org/0000-0001-9565-9665